Effect of routine diagnostic imaging for patients with musculoskeletal disorders: A meta-analysis

Yasmaine Karel
Karin Verkerk
Silvio Endenburg
Sven Metselaar
Arianne Verhagen

Published in European Journal of Internal Medicine
2016, 26(8)

ABSTRACT

Purpose

The increasing use of diagnostic imaging has led to high expenditures, unnecessary invasive procedures and/or false-positive diagnoses, without certainty that the patients actually benefit from these imaging procedures. This review explores whether diagnostic imaging leads to better patient-reported outcomes in individuals with musculoskeletal disorders.

Method

Databases were searched from inception to September 2013, together with scrutiny of selected bibliographies. Trials were eligible when: 1) a diagnostic imaging procedure was compared with any control group not getting or not receiving the results of imaging; 2) the population included individuals suffering from musculoskeletal disorders, and 3) if patient-reported outcomes were available. Primary outcome measures were pain and function. Secondary outcome measures were satisfaction and quality of life. Subgroup analysis was done for different musculoskeletal complaints and high technological medical imaging (MRI/CT).

Results

Eleven trials were eligible. The effects of diagnostic imaging were only evaluated in patients with low back pain (n=7) and knee complaints (n=4). Overall, there was a moderate level of evidence for no benefit of diagnostic imaging on all outcomes compared with controls. A significant but clinically irrelevant effect was found in favor of no (routine) imaging in low back pain patients in terms of pain severity at short [SMD 0.17 (0.04-0.31)] and long-term follow-up [SMD 0.13 (0.02-0.24)], and for overall improvement [RR 1.15 (1.03-1.28)]. Subgroup analysis did not significantly change these results.

Conclusion

These results strengthen the available evidence that routine referral to diagnostic imaging by general practitioners for patients with knee and low back pain yields little to no benefit.

Keywords

diagnostic tests, musculoskeletal/connective tissue disorders, back pain, primary care, radiology.

INTRODUCTION

For patients in whom the diagnosis remains uncertain after history taking and physical examination, general practitioners (or clinicians in general) can turn to diagnostic imaging modalities [1]. However, there has been a steady but debatable increase in the use of diagnostic imaging. For example, in the USA, between 1995 and 2005 the frequency of computed tomography (CT) has doubled and for magnetic resonance imaging (MRI) it has more than tripled [2]. The increase of diagnostic tests can lead to a false-positive diagnosis, 'pseudo' disease, or adverse effects, resulting in an unnecessary chain of events [3-6]. Imaging procedures may also lead to incidental findings, which can be found in both symptomatic and asymptomatic individuals [7, 8] indicating that diagnostic imaging findings may not always be responsible for the complaints experienced by the patient. The USA has experienced an larger number of spine surgeries due to an increase in the rate of spinal imaging [9] and others have reported increasing costs due to diagnostic imaging [10-12]. On the other hand the advancements in medical imaging techniques like MRI and other high technological medical imaging techniques can be used to replace older imaging techniques.

A previous systematic review including six randomized clinical trials (RCTs) in low back pain patients reported that immediate, routine lumbar spine imaging did not improve patient-reported outcomes [13]. Several trials have focused on patients with other musculoskeletal disorders, of which two found significant results for the effect of imaging [14-16]. Clinicians generally assume that reassurance must follow from a confident statement that no disease has been found. Nevertheless, negative test results are not always effective in reassuring patients [17]. A recent systematic review of five RCTs concluded that there is very limited evidence from current studies for the reassuring value of diagnostic tests in patients with varying complaints [18].

Although diagnostic imaging procedures are believed to influence patient care in a variety of ways, it remains unclear whether there is sufficient evidence to show that patient outcomes improve due to diagnostic imaging [13, 18]. Until now, no review has studied the effectiveness of diagnostic imaging for patients with musculoskeletal disorders other than low back pain, or has used the GRADE approach to determine the strength of the evidence. Therefore, this review aims to evaluate the role of immediate (after first consultation) diagnostic imaging procedures in patients with musculoskeletal disorders on patient-reported outcome measures (PROMs) using the GRADE approach.

METHODS

Selection criteria

RCTs were eligible when: 1) a diagnostic imaging procedure was compared with a control group not getting diagnostic imaging or not receiving results of imaging; 2) the population included individuals suffering from musculoskeletal disorders, and 3) if one of the following primary outcomes were reported: disability, pain, sick leave, quality of life, satisfaction, mental health, reassurance, or overall improvement/recovery.

Search method

Three review authors (YK,SE,SM) identified RCTs by searching the databases of MEDLINE, Cochrane, EMBASE and PubMed from inception to September 2013 (supplementary material). Relevant reference lists were also reviewed for additional citations. Two review authors (YK,KV) independently performed the study selection. Any disagreements were resolved by discussion, or with a third review author (AV), to reach consensus.

Risk of bias assessment

Two review authors (YK,KV) independently assessed the risk of bias using the Delphi list [19, 20]. In case of discrepancy, discussion was used to resolve any disagreement, or with a third review author (AV), to reach consensus. The Delphi list consists of nine items. For the present review we consider a study to have low risk bias when five or more of the items are answered with "yes"; this is supported by empirical evidence from the Cochrane Back Review Group [21].

Data extraction

Data extraction was first done by one review author (YK) using a standardized form and checked by a second author (KV), independently. When necessary, a third author (AV) resolved discrepancies. Descriptive data included study setting, country, selection criteria, population characteristics, description of intervention(s), outcomes (pain, function, quality of life, recovery and satisfaction) and follow-up. We extracted the number of participants randomized, the number of patients included in each analysis, and the means and standard deviations (SDs) of follow-up measurements.

Data analysis

Short-term follow-up was defined as being closest to 3 months and long-term follow-up as being closest to 12 months. Studies were excluded from analysis if they had insufficient data on means (or within-group differences) and SDs and the original authors could not be contacted. Pooling was done using a random effects model [22]. In case only median scores could be extracted, the median value was used as the mean and the SD was estimated from the interquartile range. For continuous outcomes the standardized mean differences (SMD) was calculated and a risk ratio (RR) for dichotomous outcomes with the accompanying 95% confidence intervals (CI). A SMD of 0-0.2 was regarded as no effect, 0.2-0.5 as a small effect, 0.5-0.8 as a moderate effect, and >0.8 as a large effect [23]. Results were considered clinically relevant when the difference between groups was ≥ 15% [24]. Wherever possible, subgroup analyses were done (separately) for different musculoskeletal complaints, study setting, and/or imaging methods (high technological imaging techniques like MRI/CT). Pooling the effects of all trials was done when heterogeneity was low ($I^2 \leq 40\%$), otherwise only the subgroup analysis was reported. Sensitivity analysis was done excluding studies with a high risk of bias, in order to control for biased results. A funnel plot evaluated publication bias only if there were ≥ 10 trials for each effect estimate; otherwise, the power of the tests would be too low to distinguish the chance from real asymmetry [25]. All analyses were conducted in Review Manager 5.2.

Strength of the evidence

The Grades of Recommendation, Assessment Development and Evaluation (GRADE) was applied to assess the overall quality of the evidence and strength of recommendations [26]. The quality of the evidence for a specific outcome was downgraded by one level for each of the factors that was encountered: 1) limitations due to study design (>25% of the included studies with a high risk of bias), 2) inconsistency of results [significant statistical heterogeneity (I^2 >40%) or inconsistent findings between the studies (≤75% of the participants report findings in the same direction)], 3) indirectness of evidence (factors affecting the generalizability of results), 4) imprecision (total number of participants <300 for each outcome), and 5) other items (e.g. reporting/publication bias, flawed design). The quality of evidence is considered to be high when RCTs with low risk of bias provide consistent, generalizable and precise results for a particular outcome [27]. Two review authors (YK,AV) scored the levels of evidence. The following levels of the quality of the evidence were applied:

- *High quality:* Further research is very unlikely to change the confidence in the estimate of the effect.
- *Moderate quality:* Further research is likely to have an important impact on confidence in the estimate of effect and may change the estimate.
- *Low quality:* Further research is very likely to have an important impact on confidence in the estimate of effect and is likely to change it.
- *Very low quality:* Great uncertainty about the estimate.

RESULTS

Results of the search and description of studies

Searching the databases resulted in 13,167 references (Figure A). After screening on title and abstract, 32 references remained. Then, screening the full-text article excluded 17 references, leaving 15 references for inclusion [11, 14-16, 28-38]. Three RCTs were published twice [15, 28, 35-38] and one trial had three different publications [11, 33, 34]. Although the DAMASK trial had 6 publications [14, 39-43] only one [14] met the inclusion criteria. One Damask publication [40] presented the trial protocol and was used for the risk of bias assessment. One of the articles [15] reported the results of two trials and was therefore regarded as two separate trials.

Finally, 10 trials were included in the analysis and their characteristics are presented in Table A.

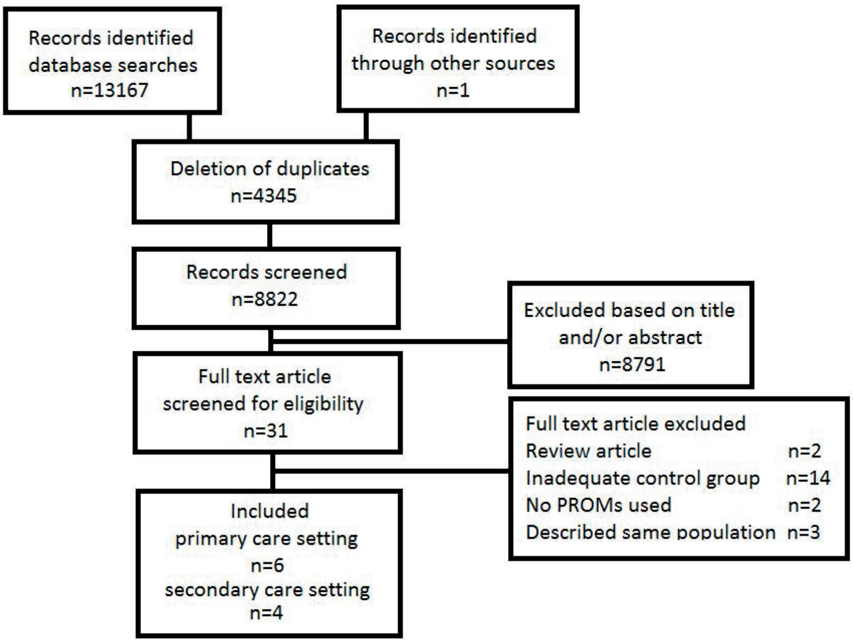


Figure A. Flow diagram

Population

The 11 trials included a total of 2,777 patients (ranging from 50-782 patients per trial); mean age ranged from 28-52 years. Seven trials included a population with acute or subacute low back pain [28-33, 35] and four trials included patients with knee complaints [14-16]. One low back pain study did not report measures of variability and was not pooled in the analysis [30].

Table A. Study Characteristics.

Study	Participants	Interventions	Outcomes	Follow-up	Notes
Ash et al and Modic et al [28,38]	Patients with acute LBP and/or radiculopathy recruited from primary care, Mean age: 43 years; 29% male, Country: US; n=246	1. MRI results provided within 48 hours (n= 131) 2. Patient and physician blinded to MRI imaging results (n= 115).	Pain (SF-36 Bodily pain (0-100)) Function (SF-36 Physical functioning (0-100)) Function (RDQ (0-24)) Pain (VAS (0-10)) absenteeism, self- efficacy scores FABQ (0-24)	2, 4, 6, 8 weeks & 6, 12 and 24 months	All patients received treatment recommendations from guidelines. Patients and physicians in control group were told the imaging results after 6 months
Brealy, DAMASK-trial [14,40]	Patients consulting their GP about continuing knee problems, Mean age: 40 years; 64% male, country: UK; n=553	1. MRI + provisional orthopedic referral (n= 279) 2. Orthopadic referral alone (n= 274)	SF-36 Bodily pain (0-100) Function (SF-36 Physical functioning (0-100)) Function (KQoL-26 (0-100)), Qol (EQ-5D (-0.59-1))	6, 12 and 24 months	112 patients in MRI group had arthroscopy after MRI versus 77 patients in the control group. Whether patients in the control group received imaging was not reported
Bryan (a) et al [15,37]*	Patients with knee complaints attending the orthopedic clinics who were considered for arthroscopy, Mean age: 36 years; 65% male, country: UK; n=118	1. MRI (n= 59) 2. No-MRI (n= 59) Trial arms were immediately listed for arthroscopy, reviewed in clinic (both before and after surgery)	QoL (EQ-5D (-0.59-1)), Pain (SF-36 Bodily pain (0-100)) Function (SF-36 Physical functioning (0-100))	6, 12 and 24 months	9 patients in MRI group did not receive an MRI, and 41% received surgery versus 71% in the control group. Surgery was mainly arthroscopy.
Bryan (b) et al [15,37])*	Patients with acute knee injuries attending the accident and emergency department, Mean age: 28 years; 70% male, country: UK; n=120	1. MRI reviewed with management plan (n= 57) 2. MRI results blinded from clinical management team (n= 63)	QoL (EQ-5D (-0.59-1)), SF-36 Bodily pain (0-100) SF-36 Physical functioning (0-100) Lysholm score (0-25)	6, 12 and 24 months	If patients in the control group still had problems after 6 weeks the protocol allowed for the MRI scan to be reviewed. 30% of MRI patients had surgery compared to 24% of no-MRI patients

Table A. Study Characteristics. (continued)

Study	Participants	Interventions	Outcomes	Follow-up	Notes
Cohen et al [29]	Patients with LBP; clinical candidates, referred by GP, for an epidural steroid injection (ESI), Mean age: 52 years; 66% male, country: US; n=132	1. Physical examination and MRI (n= 67) 2.Physical examination only (n=65)	Leg pain score (0-10), Pain (NRS (0-10)), Function (ODI (0-100%)), GPE, Medication reduction	1 month & 3 months for patients with positive results from ESI	
Deyo et al [30]	Patients with LBP presenting to the walk-in clinic of a public hospital. Mean age: 33; 76% male, country: US; n=101	1. Radiography (n= 43) 2. Education, if no improvement was seen after 3 weeks radiography was done (n= 49)	Pain (1-6) Patient satisfaction (9-27) Function (SIP (0-100)) Self-rated improvement	3 weeks and 3 months	After 3 months 29% of control group had received lumbar radiography
Djais and Kalim [31]	Patients with acute LBP attending rheumatology department, Median age: 40; 55% male, country: Indonesia, n=101	1. Radiography (n=51) 2. Usual care without radiography (n=50)	Function (RDQ (0-24)) Pain (VAS (0-10)) QoL (EQ-5D (-0.59-1)) Overall improvement	3 weeks	Drop-out rate 28%
Gilbert et al [32]	Patients with lumbar spine disorder referred by their GP to an orthopaedic specialist or neurosurgeon, Mean age: 43; 49% male, country: UK; n=782	1. CT or MRI (n=393) 2. Delayed imaging (no MRI or CT unless a clear clinical indication developed) (n=389)	Function (ALBP (0-100)) EQ-5D (-0.59-1)) Pain (SF-36 Bodily Pain (0-100)) Function (SF-36 Physical Functioning (0-100))	8 and 24 months	After 24 months 30% of the control group had received imaging
Kendrick et al and Miller et al [11,33,34]	Patients with LBP consulting the GP, Median age: 39; 87% male, country: UK; n=421	1. Radiography (n= 210) 2. Usual care without radiography (n= 211)	Function (RDQ (0-24)) Pain (VAS (0-10)) QoL (EQ-5D (-0.59-1)) Satisfaction (9-27)	3 and 9 months	12% of patients in the control group had radiography Over 78% drop-out on satisfaction

Table A. Study Characteristics. (continued)

Study	Participants	Interventions	Outcomes	Follow-up	Notes
Kerry et al [35, 36]	Patients consulting their general practitioner with LBP, Mean age: 44; 51% male, country: UK; n=153	1. Referral for radiography (n= 73) 2. Usual care without referral for lumbar radiography (n= 80)	Function (RDQ (0-24)) HADS Pain (SF-36 Bodily pain (0-100)) SF-36 Physical functioning (0-100) QoL (EQ-5D (0-100))	6 weeks and 1 year	Baseline differences, the intervention group had a longer duration of complaints and a higher score on the SF-36. Drop-out >20% After 12 months 14% of the control group had received radiography
Patel et al [16]	Patients with acute knee complaint due to a twisting injury Mean age: 30; 72% male, country: UK; n=50	1. MRI within 2 weeks (n= 25) 2. Conventional management with physiotherapy (n= 25)	Pain (VAS (0-10)), Function (Activity limitation (0-10)) Satisfaction (0-10)	2 weeks and 3 months	1 person in the control group opted for an MRI

LBP Low Back Pain, MRI Magnetic Resonance Imaging, NRS Numeric Rating Scale, GPE Global Perceived Effect, SF-36 Short Form 36 item, Kqol-26 Knee Quality of Life 26 item, QoL Quality of Life, EQ-5D EuroQol 5 dimensions, SIP Sickness Impact Profile, VAS Visual analog scale, ALBP Aberdeen Low Back Pain score, FABQ Fear Avoidance Beliefs Questionnaire GP General Practitioner, HADS Hospital Anxiety and Depression Scale, US United States, UK United Kingdom

* reported data extracted from figures without the actual numbers

Seven trials were performed in the UK [14-16, 32, 33, 35], three in the USA [28-30] and one in Indonesia [31]. The study setting was either primary [14, 28, 29, 32, 33, 35] or secondary health care [15, 16, 30, 31]. Four trials specified the duration of complaints in their inclusion criteria; this ranged from ≤ 1 week to 12 weeks [11, 16, 28, 31].

Interventions

Six trials used MRI as the diagnostic imaging procedure [14-16, 28, 29], one of these used either CT or MRI [32] and four trials used radiography [30, 31, 33, 35]. Five trials compared immediate or early imaging with usual care [16, 30-33]. Four had a control group that could receive imaging based on the usual care trajectory, and two of these trials reported a waiting time for imaging ranging from 29 days [15] to 12 weeks [14]. One trial provided MRI results to the intervention group within 48 h while the control group was blinded to the MRI results [28]. Two trials [15] compared arthroscopy with MRI and arthroscopy alone and in one trial [29] all patients received an epidural steroid injection either based on history and physical examination, or on clinical findings and imaging results.

Five trials [14, 30, 32, 33, 35] reported the percentage of patients receiving imaging in the control group (ranging from 2-30%) as part of usual care.

Outcome measures

All trials assessed both pain and function (Table 1). Five trials examined pain with the Bodily Pain score of the Short Form 36 (SF-36). Four trials reported pain with a visual analogue scale (VAS) or the numerical rating scale (NRS) and another trial rated pain on a six-point scale. To assess function, both generic and disease-specific measurement instruments were used. Disease-specific measurement instruments were the Roland Disability Index, the Aberdeen Low Back Pain Score, the Oswestery Disability Index, the Lysholm score, and the Knee Quality of Life Questionnaire.

Two trials reported median scores and interquartile ranges [31, 33]. For pain and function, all outcome measures were continuous. In five trials overall improvement was measured as dichotomous. Two of these five trials reported satisfaction on an ordinal scale and two on a continuous scale, of which one had a 78% dropout rate on this outcome and was excluded from the analysis (fatal flaw) [33]. Only the results for dichotomous outcomes were pooled.

For two trials we contacted the authors for additional information. For one [33] of these trials we received information from the author about a systematic review including this trial [13, 33]. Another trial did not report data to impute SDs [30]; unfortunately, we did not receive any response from these authors. Because one article [15] only reported data in figures, the data were estimated from these figures.

Risk of bias assessment

Six trials (55%) were considered to have low risk of bias [14, 15, 28, 29, 33]. Overall, risk of bias was threatened by the inability to blind patients (n=10), care providers (n=11) or outcome assessors (n=10), and by the absence of an intention-to-treat analysis (n=8). Concealment of randomization was not adequately reported in three trials. The results of the risk of bias assessment are presented in (supplementary material).

Effects of imaging

All effects estimates are described in the summary of findings (Table B). GRADE scoring is reported for short and long-term follow-up and (separately) for low back pain and knee studies. Only subgroup results are reported when heterogeneity was high for the overall effect estimate.

Pain. Figure B.1-2. shows the improvement in pain on short and long-term follow-up. Pooling the studies with low back pain patients resulted in a significant effect in favor of no imaging on the short [SMD 0.17 (95%CI: 0.04-0.31)] and long term [SMD 0.13 (95%CI: 0.02-0.24)] but the effect size was below 0.2, while the trials with patients with knee complaints found no difference on the long term [SMD 0.02 (95%CI: -0.14-0.18)]. In the short-term analysis only one study with knee complaints had available results on pain; these results indicated a non-significant effect in favor of imaging (Figure B.1). Heterogeneity was small (I^2=39%) at short-term follow-up and not present at long-term follow-up. When all trials were pooled, no significant and clinically relevant differences were found on the short term [SMD 0.10 (95%CI: -0.08-0.29)]. On long-term follow-up data showed borderline significant results in favor of no imaging [SMD 0.09 (95%CI: 0.00-0.18)] but the effect size remained below 0.2.

In the short-term analysis there were four studies and in the long-term analysis five studies with a primary care population. Effects sizes for both the short term [SMD 0.15 (95%CI: 0.01-0.30)] and long term [SMD 0.11 (95%CI: 0.01-0.20)] resulted in borderline significant effects in favor of no imaging but the effect size was below 0.20.

Pooling only the trials using radiography (n=3) as imaging method resulted in a significant effect in favor of no imaging but a SMD below 0.2 [SMD 0.15 (95%CI: 0.03-0.26)], whereas pooling the trials with MRI (n=8) found no difference [SMD 0.07 (95%CI: -0.05-0.18)] (data not shown).

Overall, we found moderate level of evidence (downgraded based on limitations in study design) for a small clinically irrelevant effect on pain in favor of no imaging on the long term, especially for the low back pain trials and trials using radiography.

Table B. Summary of findings table.

Patient recovery after diagnostic imaging in patients with musculoskeletal disorders

Population: patients with musculoskeletal complaints
Intervention: Diagnostic imaging (radiography/MRI)
Setting: Primary care/secondary care emergency department/orthopaedic department.
Comparison: Usual care/not getting immediate imaging/not receiving results of imaging

Outcomes	Standardized mean difference (95% Confidence Interval)	Number of participants (studies)	Quality of evidence (GRADE score)	Comments
Pain short-term	0.10 (-0.08; 0.29)	890 (6 studies)	Moderate level [2]	Two studies the median was used as mean and SD was calculated from IQR
LBP studies	0.17 (0.04; 0.31)	814 (5 studies)	Low level [1, 2]	
Knee pain studies	-0.51 (-1.10 to 0.07)	46 (1 study)	Very low level [1, 2, 3]	
Primary care studies	0.15 (0.01 to 0.30)	768 (3 studies)	Moderate level [2]	
Pain long term	0.09 (0.00; 0.18)	1875 (7 studies)	Moderate level [2]	One study only reported 24 months follow-up, another only figures and for 1 study the median was used as mean and SD was calculated from IQR
LBP studies	0.13 (0.02; 0.24)	1281 (4 studies)	Moderate level [2]	
Knee pain studies	0.02 (-0.14; 0.18)	594 (3 studies)	High level	
Primary care studies	0.11 (0.01 to 0.20)	1752 (5 studies)	Moderate level [2]	
Function gen. short term	-0.21 (-0.55; 0.12)	348 (3 studies)	Low level [1, 2]	Two studies the median was used as mean and SD was calculated from IQR
LBP studies	-0.12 (-0.49; 0.25)	302 (2 studies)	Low level [1, 2]	
Knee pain studies	-0.57 (-1.16; 0.02)	46 (1 study)	Very low level [1, 2, 3]	
Primary care studies	-0.12 (-0.49 to 0.25)	302 (2 studies)	Low level [1, 2]	
Function gen. long term	0.08 (-0.05; 0.20)	1481 (6 studies)	Moderate level [2]	One study only reported 24months follow up, another only figures and for 1 study the median was used as mean and SD was calculated from IQR
LBP studies	0.10 (-0.03; 0.23)	887 (3 studies)	Moderate level [2]	
Knee pain studies	-0.07 (-0.44; 0.31)	594 (3 studies)	Moderate level [1]	
Primary care studies	0.13 (0.02 to 0.24)	1358 (4 studies)	Moderate level [1]	
Function spec. short-term	0.11 (-0.04; 0.27)	844 (5 studies)	Moderate level [2]	Two studies the median was used as mean and SD was calculated from IQR
Primary care studies	0.09 (-0.06 to 0.23)	678 (4 studies)	Moderate level [2]	
Function spec. long-term	0.01 (-0.34; 0.41)	1281 (4 studies)	Low level [1, 2]	One study only reported 24months follow up, another only figures and for 1 study the median was used as mean and SD was calculated from IQR
tLBP studies	0.04 (-0.38; 0.45)	525 (2 studies)	Moderate level [1]	
Knee pain studies				

Table B. Summary of findings table. (continued)

Outcomes	Standardized mean difference (95% Confidence Interval)			Number of participants (studies)	Quality of evidence (GRADE score)	Comments
QoL short-term LBP studies	-0.00 (-0.07; 0.06)			202 (2 studies)	Very low level [1, 2, 3]	
QoL long-term	0.01 (-0.12; 0.10)			1270 (7studies)	Moderate level [2]	
LBP studies	0.03 (-0.09; 0.14)			1143 (5 studies)	Moderate level [2]	
Knee studies	-0.18 (-0.54; 0.18)			123 (2 studies)	Moderate level [3]	
Primary care studies	-0.03 (-0.14 to 0.09)			1143 (3 studies)	Low level [1, 2]	
	Assumed mean score/ comparative risk (95% CI)		Relative Effect (95% CI)	Number of participants (studies)	Quality of evidence (GRADE)	
	Imaging	No imaging				
Overall improvement LBP studies	128 of 343 recovered	151 of 335 recovered	RR 1.15 (1.03; 1.28)	855 (5 studies)	Moderate level [2]	
Satisfaction	66 of 150 satisfied	73 of 152 satisfied	RR 1.03 (0.85; 1.24)	302 (2 studies)	Low level [1, 2]	

[1] Downgraded because of inconsistency

[2] Downgraded because of limitations in study design

[3] Downgraded because of imprecision

SD standard deviation, CI confidence interval, IQR interquartile range, LBP Low Back Pain, Gen Generic, Spec Specific, QoL Quality of Life

Figure B.1.

Pain intensity short-term

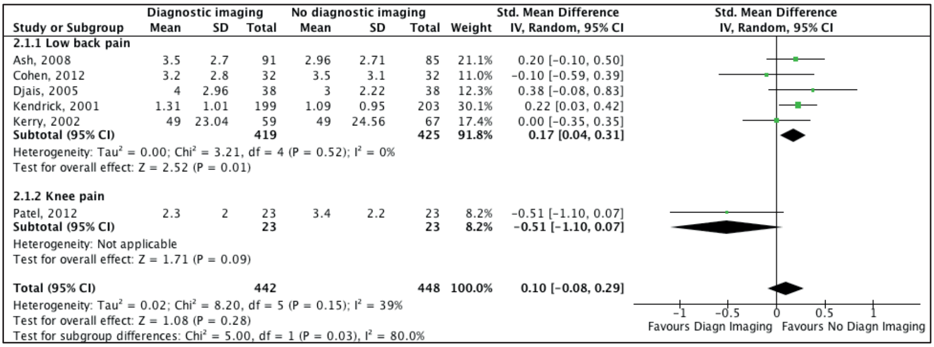


Figure B.2.

Pain intensity long-term

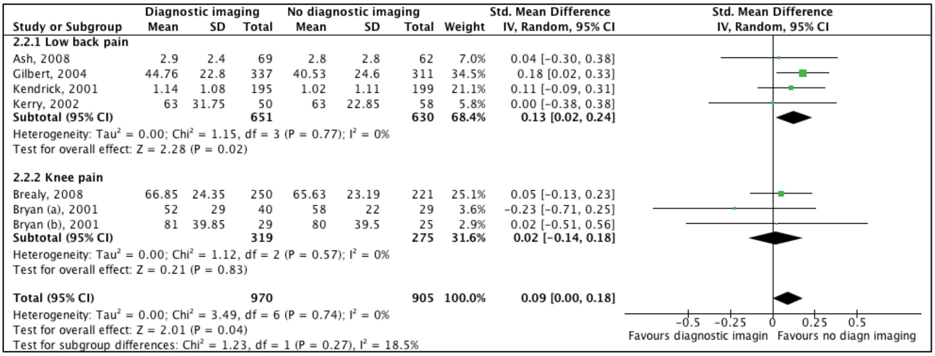


Figure B. Pain intensity long-term and short-term

Function. Figure C.1-4. shows the improvement in function measured with generic and specific measurement instruments for short and long-term follow-up.

Heterogeneity was present at short-term outcome (I^2=55%) and small for long-term outcome. Subgroup analysis for patients with low back pain had non-significant differences at short term [SMD -.021 (95%CI: -0.55-012)] and long term [SMD 0.10 (95%CI: -0.03-0.23)]. Trials with knee complaints were only available for the long-term results and showed a non-significant effect in favor of imaging [SMD -0.07 (95%CI: -0.44-0.31)]. The overall effect estimate for knee and low back pain studies combined at long term found no effects and were not significant [SMD 0.08 (95%CI: -0.05-0.20)].

In the short-term analysis there were two studies and in the long-term analysis four studies with a primary care population. Effects sizes for the short term [SMD -0.12 (95%CI: -0.49-0.25)] were not significant. Long-term analysis resulted in a small borderline significant effect [SMD 0.13 (95%CI: 0.02-0.24)] in favor of no imaging but a SMD below 0.2.

Excluding the only trial using radiography as a method of imaging resulted in a non-significant effect estimate in the MRI subgroup [SMD -0.08 (95%CI: -0.27-0.11)] (data not shown).

We found low level evidence (downgraded based on limitations in study design and inconsistency) that there is no difference on the short term and moderate level of evidence (downgraded based on study design and inconsistency) on the long term for function measured with generic measurement instruments.

Figure C3-4 shows improvement in function with disease-specific instruments. Heterogeneity was very small for the short term (I^2=16%) because no trials with knee complaints were available. Substantial heterogeneity was present at long-term follow-up (I^2=70%). Both outcome measures are reported per subgroup. Subgroup analysis for low back pain trials resulted in a non-significant effect on the short term [SMD 0.11 (95%CI: -0.04-0.27)] and long term [SMD 0.01 (95%CI: -0.23-0.25)].

The short-term analysis included four studies with a primary care population; pooling these studies did not significantly alter the effect size [SMD 0.09 (95%CI: -0.06-0.23)]. All studies in the long-term analysis were primary care populations.

Pooling studies with knee complaints resulted in a non-significant effect [SMD 0.04 (95% CI -0.38; 0.45)].

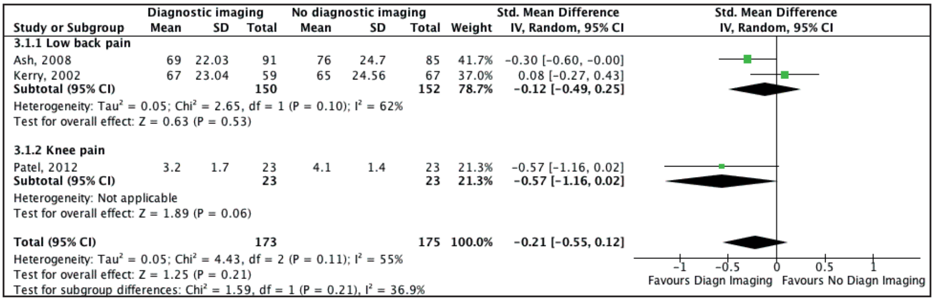
No differences were found [SMD 0.01 (95%CI: -0.19-0.21)] when analysing trials using MRI (n=6). Pooling trials using radiography (n=3) resulted in a borderline significant difference [SMD 0.13 (95%CI: -0.00-0.25)] in favor of the no imaging group (data not shown) but the SMD was below 0.2. Separate analyses for primary care studies were not possible because of the small number of available studies.

We found moderate level of evidence (downgraded based on limitations in study design) for no differences between both groups at short-term follow-up for patients with low back pain, and low level of evidence (downgraded based on limitations in study design and inconsistency) that there is no difference between imaging and no imaging for disease-specific function at long-term follow-up, irrespective of the subgroups. Subgroup analysis found a small borderline significant effect in favor of the no imaging group in trials using radiography.

Satisfaction. Moderate heterogeneity was present (I^{2}=44%). Because of the limited number of trials the short and long-term results were combined (data not shown). Overall, we found a low level of evidence for no differences (downgraded based on limitations in study design and inconsistency) between the groups [RR 1.03 (95%CI: 0.85-1.24)].

Figure C.1.

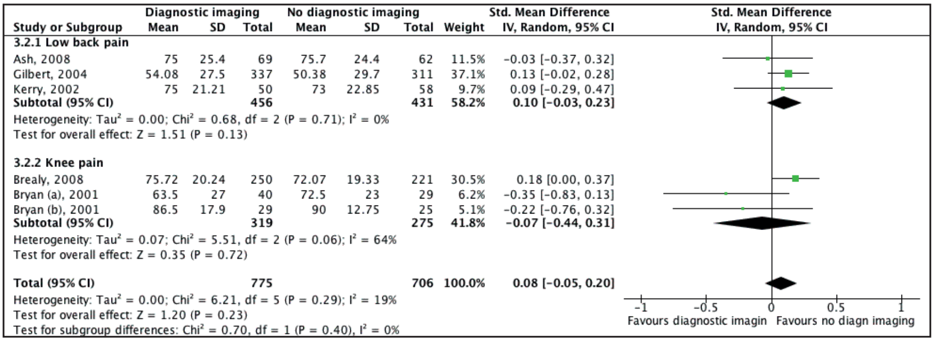
Function measured with generic instruments short-term



Study or Subgroup	Diagnostic imaging Mean	SD	Total	No diagnostic imaging Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
3.1.1 Low back pain								
Ash, 2008	69	22.03	91	76	24.7	85	41.7%	−0.30 [−0.60, −0.00]
Kerry, 2002	67	23.04	59	65	24.56	67	37.0%	0.08 [−0.27, 0.43]
Subtotal (95% CI)			**150**			**152**	**78.7%**	**−0.12 [−0.49, 0.25]**
Heterogeneity: Tau² = 0.05; Chi² = 2.65, df = 1 (P = 0.10); I² = 62%								
Test for overall effect: Z = 0.63 (P = 0.53)								
3.1.2 Knee pain								
Patel, 2012	3.2	1.7	23	4.1	1.4	23	21.3%	−0.57 [−1.16, 0.02]
Subtotal (95% CI)			**23**			**23**	**21.3%**	**−0.57 [−1.16, 0.02]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 1.89 (P = 0.06)								
Total (95% CI)			**173**			**175**	**100.0%**	**−0.21 [−0.55, 0.12]**
Heterogeneity: Tau² = 0.05; Chi² = 4.43, df = 2 (P = 0.11); I² = 55%								
Test for overall effect: Z = 1.25 (P = 0.21)								
Test for subgroup differences: Chi² = 1.59, df = 1 (P = 0.21), I² = 36.9%								

Figure C.2.

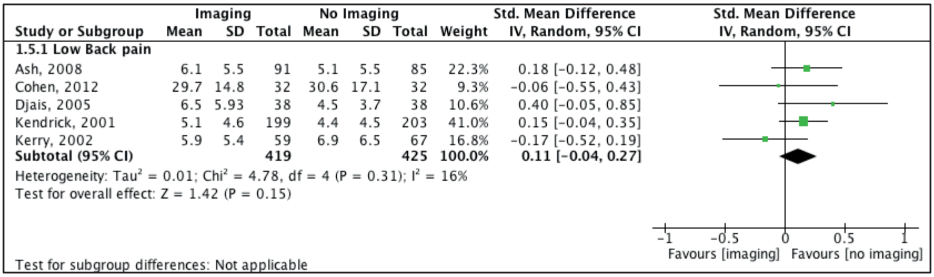
Function measured with generic instruments long-term



Study or Subgroup	Diagnostic imaging Mean	SD	Total	No diagnostic imaging Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
3.2.1 Low back pain								
Ash, 2008	75	25.4	69	75.7	24.4	62	11.5%	−0.03 [−0.37, 0.32]
Gilbert, 2004	54.08	27.5	337	50.38	29.7	311	37.1%	0.13 [−0.02, 0.28]
Kerry, 2002	75	21.21	50	73	22.85	58	9.7%	0.09 [−0.29, 0.47]
Subtotal (95% CI)			**456**			**431**	**58.2%**	**0.10 [−0.03, 0.23]**
Heterogeneity: Tau² = 0.00; Chi² = 0.68, df = 2 (P = 0.71); I² = 0%								
Test for overall effect: Z = 1.51 (P = 0.13)								
3.2.2 Knee pain								
Brealy, 2008	75.72	20.24	250	72.07	19.33	221	30.5%	0.18 [0.00, 0.37]
Bryan (a), 2001	63.5	27	40	72.5	23	29	6.2%	−0.35 [−0.83, 0.13]
Bryan (b), 2001	86.5	17.9	29	90	12.75	25	5.1%	−0.22 [−0.76, 0.32]
Subtotal (95% CI)			**319**			**275**	**41.8%**	**−0.07 [−0.44, 0.31]**
Heterogeneity: Tau² = 0.07; Chi² = 5.51, df = 2 (P = 0.06); I² = 64%								
Test for overall effect: Z = 0.35 (P = 0.72)								
Total (95% CI)			**775**			**706**	**100.0%**	**0.08 [−0.05, 0.20]**
Heterogeneity: Tau² = 0.00; Chi² = 6.21, df = 5 (P = 0.29); I² = 19%								
Test for overall effect: Z = 1.20 (P = 0.23)								
Test for subgroup differences: Chi² = 0.70, df = 1 (P = 0.40), I² = 0%								

Figure C.3.

Function measured with disease specific instruments short-term



Study or Subgroup	Imaging Mean	SD	Total	No Imaging Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.5.1 Low Back pain								
Ash, 2008	6.1	5.5	91	5.1	5.5	85	22.3%	0.18 [−0.12, 0.48]
Cohen, 2012	29.7	14.8	32	30.6	17.1	32	9.3%	−0.06 [−0.55, 0.43]
Djais, 2005	6.5	5.93	38	4.5	3.7	38	10.6%	0.40 [−0.05, 0.85]
Kendrick, 2001	5.1	4.6	199	4.4	4.5	203	41.0%	0.15 [−0.04, 0.35]
Kerry, 2002	5.9	5.4	59	6.9	6.5	67	16.8%	−0.17 [−0.52, 0.19]
Subtotal (95% CI)			**419**			**425**	**100.0%**	**0.11 [−0.04, 0.27]**
Heterogeneity: Tau² = 0.01; Chi² = 4.78, df = 4 (P = 0.31); I² = 16%								
Test for overall effect: Z = 1.42 (P = 0.15)								
Test for subgroup differences: Not applicable								

Figure C.4.

Function measured with diseases specific instruments long-term

Figure C. Function short-term & long-term

Study or Subgroup	Imaging Mean	Imaging SD	Imaging Total	No Imaging Mean	No Imaging SD	No Imaging Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.6.1 Low Back pain								
Ash, 2008	4.9	5.3	69	4.2	5.3	62	20.7%	0.13 [−0.21, 0.47]
Gilbert, 2004	32.67	17.4	337	36.98	19.8	311	31.4%	−0.23 [−0.39, −0.08]
Kendrick, 2001	4.4	4.9	195	3.6	4.5	199	29.0%	0.17 [−0.03, 0.37]
Kerry, 2002	4.5	5.7	50	4.3	5.3	58	19.0%	0.04 [−0.34, 0.41]
Subtotal (95% CI)			**651**			**630**	**100.0%**	**0.01 [−0.23, 0.25]**
Heterogeneity: $Tau^2 = 0.04$; $Chi^2 = 11.31$, df = 3 (P = 0.01); $I^2 = 73\%$								
Test for overall effect: Z = 0.09 (P = 0.93)								
1.6.2 Knee pain								
Brealy, 2008	75.72	20.24	250	72.07	19.33	221	66.7%	0.18 [0.00, 0.37]
Bryan (b), 2001	60	7.5	29	62	7.5	25	33.3%	−0.26 [−0.80, 0.27]
Subtotal (95% CI)			**279**			**246**	**100.0%**	**0.04 [−0.38, 0.45]**
Heterogeneity: $Tau^2 = 0.06$; $Chi^2 = 2.38$, df = 1 (P = 0.12); $I^2 = 58\%$								
Test for overall effect: Z = 0.17 (P = 0.87)								
Test for subgroup differences: $Chi^2 = 0.01$, df = 1 (P = 0.92), $I^2 = 0\%$								

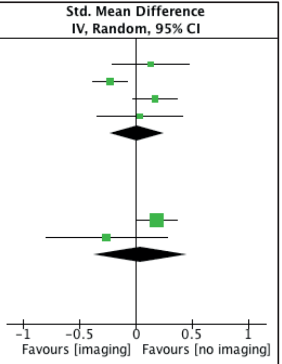


Figure C.1-4.

Function short-term & long-term

Figure C. Function short-term & long-term *(continued)*

Quality of Life. Figure D shows the results of 'quality of life' for the short and long-term follow-up.

Substantial heterogeneity was present at short-term (I^2=86%) but not at long-term follow-up (figure D). For the short-term pooled effect estimate, only two low back pain studies were available [SMD -0.07 (95%CI: -0.83-0.68)]. Subgroup analysis for the long-term effect resulted in slightly different non-significant effects between knee [SMD 0.18 (95%CI: -0.18-0.54)] and low back pain studies [SMD -0.03 (95%CI: -0.14-0.09)]. The overall effect at the long term showed no difference for knee and low back pain studies combined [SMD 0.01 (95%CI: -0.10-0.12)]. Pooling the studies performed in primary care did not significantly alter the effect size [SMD -0.03 (95%CI: -0.14-0.09)]. Overall, low level of evidence (downgraded because of limitations in study design and inconsistency) was found for no difference concerning quality of life for patients with knee pain and with low back pain at short-term follow-up and moderate level of evidence at long-term follow-up.

Overall improvement. Figure E shows the results of 'overall improvement'.

Short and long-term results were combined due to the limited number of trials reporting overall improvement. No studies with knee pain presented results for overall improvement. Heterogeneity was not present. Overall improvement showed a significant but clinically irrelevant result in favor of the no imaging group (RR 1.15, 95%CI: 1.03-1.28). Sensitivity analysis showed that excluding two trials [30, 31] with high risk of bias did not change the results (RR 1.13, 95%CI: 1.01-1.27).

Four studies were performed in primary care; pooling these studies did not alter the results (RR 1.15, 95%CI: 1.03-1.28).

Figure D.1.

Quality of Life short-term

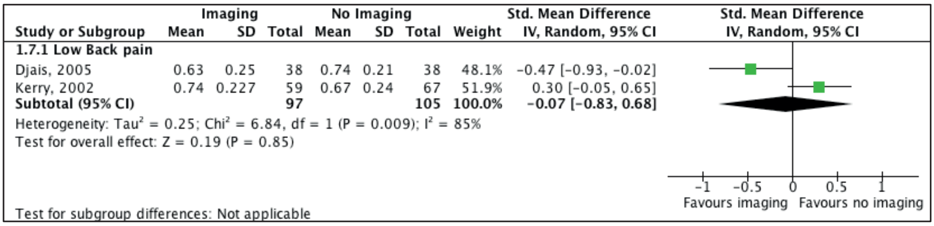


Figure D.2.

Quality of life long-term

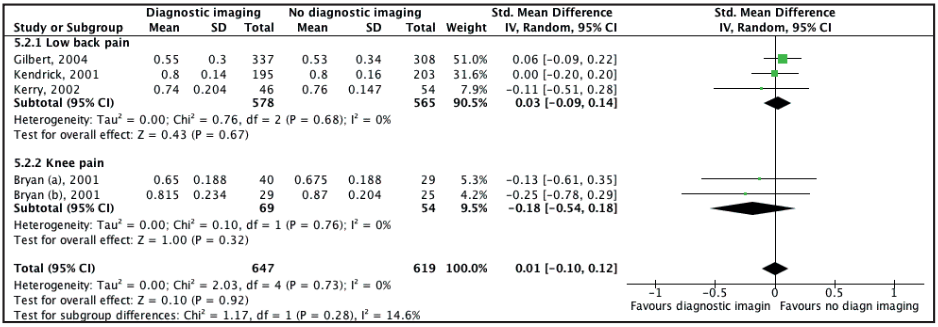


Figure D. Quality of life short- & long-term

We found a moderate level of evidence (downgraded because of limitations in study design) for a small effect in favor of no imaging concerning overall improvement for patients with low back pain.

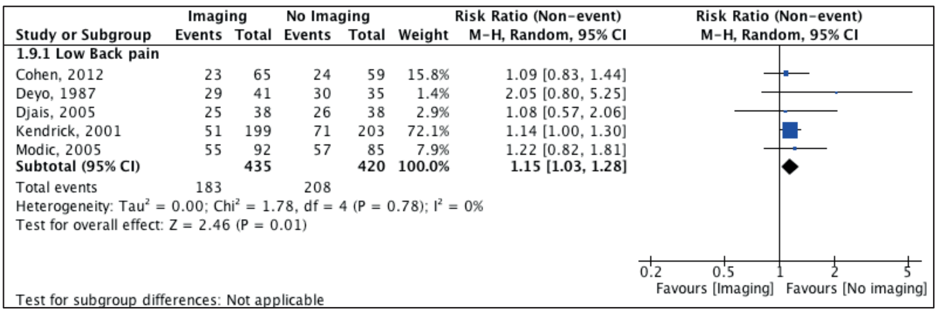


Figure E. Overall improvement long term

DISCUSSION

Overall, our results showed that early imaging strategies do not improve patient-reported outcomes (PROMs) in patients with low back pain or knee complaints. Small differences were found between these subgroups in pain, function and quality of life, in which the low back pain subgroup usually had larger effect sizes in favor of "no imaging". Notably, more trials concerning low back pain were available. Subgroup analysis in low back pain patients led to small significant effect in favor of no routine imaging. The majority of imaging tests used in low back pain show an absence of abnormality; however, this may not reassure patients and can lead to possible negative effects of imaging.

Strengths and limitations

An important strength of the present review is the sensitive search strategy applied to reduce the chance of missing relevant studies and thereby reducing publication bias. In a search strategy, defining 'diagnostic imaging' as the intervention appeared to be somewhat difficult. However, because of the sensitive search strategy (including all possible synonyms) it is unlikely that relevant trials were missed.

We aimed to include patients with all kinds of musculoskeletal disorders, acute as well as subacute, or chronic complaints. Subgroup analysis was possible for patients with knee or low back pain, and no significant differences between these subgroups were found (except for pain). The populations with knee complaints were mostly traumatic or acute knee complaints, probably having a different clinical course than that of low back pain; this might explain the differences found regarding pain. Subgroup analysis for primary care studies did not alter the results.

Another source of heterogeneity could arise from the different types of imaging modalities used in the trials. Subgroup analysis showed a significant difference between imaging modalities for radiography on pain and 'borderline' significant difference on function measured with disease-specific instruments; however, this might be based on chance and here no differences between knee and low back pain were made.

Overall, 45% of the trials scored high risk of bias. Differences in study design could have caused heterogeneity. Only one trial [29] was able to blind their patients for the allocated intervention. Given the nature of the intervention and the clinical setting of most of the trials, blinding of patients was difficult and might have caused some bias [49, 50].

Standardization of treatment after imaging was underreported and could clearly account for bias in the study results. All studies reported that some sort of treatment was provided after imaging. Treatment might have influenced the outcome of interest by increasing or decreasing the contrast between the groups. Two trials [14, 15] even reported having arthroscopy of the knee after imaging, and another study [29] used imaging in the experimental group who received an epidural steroid injection, thereby biasing the effect of imaging on the outcome. Although these co-interventions will affect outcome, the effect of imaging is to influence treatment decisions such as these. In the future, these treatment decisions will differ between intervention and control groups. Furthermore, usual care was hardly described and might differ between different countries. In several trials the control group could also receive diagnostic imaging as part of usual care, thereby decreasing the contrast between the intervention and control treatment. Although including trials performed in a clinical setting can increase generalizability, it can also affect the validity and reliability of the results.

Because all trials used valid PROMs, pooling of results was possible. Disease-specific and generic measurement instruments for function were pooled separately. Generic instruments tended to be in favor of imaging at short term, while the results generated by the disease-specific instruments tended to find no differences. In contrast, disease-specific instruments might be more responsive compared to generic instruments, or the results might be attributed to measurement error [51, 52]. All outcome measures were patient reported outcome measures. Whether treatment regime changed due to the "intervention" is unknown.

The fact that all trials excluded patients suspected of having a serious underlying condition shows the effect for diagnostic imaging in patients were its still uncertain whether it may have a favorable effect. All trials, but one [18] (who excluded one patient with malignancy), did not report finding any serious underlying conditions.

In clinical trials the patient population is selected using strict selection criteria, which also hampers generalizability of the results. In the present review, caution is needed when drawing conclusions because of the small number of studies in the subgroup analysis, the considerable amount of risk of bias, and the diversity of the study settings.

Comparison with existing literature

Results from our review are comparable with those from an earlier review [13], although we found a clearer tendency towards benefit of no imaging for low back pain patients. Another review [18] studied the effect of diagnostic imaging on reassurance, and included five trials with populations also having chest pain or headache. The latter review included two trials with musculoskeletal complaints, which were also included in our review.

Implications

In patients with a musculoskeletal disorder, imaging did not lead to better PROMs. On the contrary, some results showed a tendency towards better outcomes after no routine imaging. Other factors, like exposure to radiation, increasing costs, and use of unnecessary invasive procedures, might also influence the clinical benefit for patients.

Imaging has its place in health care where serious conditions are suspected or when surgery is considered. The natural history of low back pain is benign, as 90% of patients recover within 6 weeks, and resolves with little intervention without knowing the anatomic diagnosis [53]. In the first 6 weeks diagnostic imaging should be used in the presence of red flags (smoking, age, history of cancer, diabetes, drug abuse, chronic NSAID, unnatural course of pain, night pain or symptoms of cauda iquina) [54, 55]. Patient with complaints longer than 6 weeks, diagnostic imaging does not necessarily disclose clear pathologic diagnoses [56]. Degenerative findings are common and whether these findings can attribute to the complaints remains unknown. It seems that the results might be limited by our current inability to understand this complex multifactorial condition and future research should focus on the ability to diagnose the condition.

Patients with knee complaints reported 25% recovery after 3 months and 44% after 12 months in primary care [57]. Urgent referral to a specialist is necessary when there are signs of fracture, acute locked knee or severe pain after patella dislocation at the initial consultation that is likely to be attributed to a trauma [58]. Several clinical decision rules are validated that identify patients with a high risk of fracture [59-61]. Diagnostic imaging can also be helpful in establishing the correct diagnosis in non-traumatic knee complaints. In order to prevent excessive imaging, especially the number of images without pathology, patients should be managed conservatively and imaging should be considered when patients show no improvement [58]. According to the America College of Radiography the initial imaging study for non-traumatic knee pain should be radiography [62]. MRI is needed to further examine intra-articular abnormalities (lesions of ligaments, tendons, bone, cartilage and menisci) [63-65].

Future research should focus on trials with low risk of bias, paying special attention to standardization and blinding of trial participants. Also, future trials should try to 'prevent' patients in the usual care group from receiving any type of imaging. Furthermore, reporting the effect of clinical decisions (e.g. the number of patients having surgery or therapy) in the long term is required to study the clinical impact of imaging.

CONCLUSIONS

Routine diagnostic imaging in patients with low back pain or knee complaints did not change the outcome for pain, function, quality of life, recovery nor satisfaction. In patients with low back pain routine imaging may even cause some harm. Our results indicate that it is unlikely that use of routine diagnostic imaging in all patients leads to better patient-reported outcome measures. Imaging has its place in health care where serious conditions are suspected or when surgery is considered. Diagnostic imaging can be considered in patients with low back pain to rule out a serious underlying condition in the presence of red flags and in subactute/chronic low back pain patients who show no improvement. Clinical decision rules should be used by clinicians in patients with traumatic knee complaints. In non-traumatic knee complaints diagnostic imaging should be used if conservative treatment fails.

Caution is required when drawing conclusions, due to the small number of studies with heterogeneity in patient populations and the presence of risk of bias in a considerable percentage of the studies.

Learning points

Evidence from trials comparing routine diagnostic imaging with usual care or no imaging has yielded conflicting results.

Results from this review show small significant effects on pain and overall improvement, especially for patients with low back pain, not in favour of imaging. No different effects after receiving diagnostic imaging were found among patients with knee pain.

REFERENCES

1. Bussières, A., Peterson, C., Taylor, J.A. (2007). Diagnostic imaging practice guidelines for musculoskeletal complaints in adults an evidence-based approach: introduction. *Journal of Manipulative and Physiological Therapeutics*, 30(9), 617-83.
2. Baker, L., Atlas, S.W., Afendulis, C.C. (2008). Expanded use of imaging technology and the challenge of measuring value. *Health Affairs*, 27(6), 1467-78.
3. Fisher, E., Welch, H.G. (1999). Avoiding the unintended consequences of growth in medical care: how might more be worse? *Journal of the American Medical Association*, 281(5), 446-53.
4. Moyniham, R., Doust, J., Henry, D. (2012). Preventing overdiagnosis: how to stop harming the healthy. *British Medical Journal*, 344, e3502.
5. Wassenaar, M., van Rijn, R.M., van Tulder, M.W., Verhagen, A.P., van der Windt, D.A.W.M., Koes, B.W., de Boer, M.R., Ginai, A.Z., Ostelo, R.W.J.G. (2012). Magnetic resonance imaging for diagnosing lumbar spinal pathology in adult patients with low back pain or sciatica: a diagnostic review. *European Spine Journal*, 21(2), 220-227.
6. van Rijn, R., Wassenaar, M., Verhagen, A.P., Ostelo, R.W., Ginai, A.Z., de Boer, M.R., van Tulder, M.W., Koes, B.W. (2012). Computed tomography for the diagnosis of lumbar spinal pathology in adult patients with low back pain or sciatica: a diagnostic review. *European Spine Journal*, 21(2), 228-239.
7. Boden, S., Davis, D.O., Dina, T.S., Patronas, N.J., Wiesel, S.W. (1990). Abnormal magnetic-resonance scans of the lumbar spine in asymptomatic subjects: A prospective investigation. *The Journal of Bone and Joint Surgery American Volume*, 72(3), 403-8.
8. Jensen, M., Brant-Zawadzki, M.N., Obuchowski, N., Modic, M.T., Malkasian, D., Ross, J.S. (1994). Magnetic resonance imaging of the lumbar spine in people without back pain. *The New England Journal of Medicine*, 331(2), 69-73.
9. Lurie, J., Birkmeyer, N.J., Wernstein, J.N. (2003). Rates of advanced spinal imaging and spine surgery. *Spine*, 28(6), 616-20.
10. Maniadakis, N., Gray, A. (2000). The economic burden of back pain in the UK. *Pain*, 84(1), 95-103.
11. Miller, P., Kendrick, D., Bentley, E., Fielding, K. (2002). Cost-effectiveness of lumbar spine radiography in primary care patients with low back pain. *Spine*, 27(20), 2291-7.
12. Oei, E., Nikken, J.J., Ginai, A.Z., Krestin, G.P., Verhaar, J.A., van Vugt, A.B., Hunink, M.G. (2009). Costs and effectiveness of a brief MRI examination of patients with acute knee injury. *European Radiology*, 19(2), 409-18.
13. Chou, R., Fu, R., Carrino, J.A., Deyo, R.A. (2009). Imaging strategies for low-back pain: systematic review and meta-analysis. *Lancet*, 373(9662), 463-72.
14. Brealy, S., DAMASK (Direct Access to Magnetic Resonance Imaging: Assessment for Suspect Knees) Trial Team. (2008). Effectiveness of GP access to magnetic resonance imaging of the knee: a randomized trial. *British Journal of General Practice*, 58(556), e1-e9.
15. Bryan, S., Weatherburn, G., Bungay, H., Hatrick, C., Salas, C., Parry, D., Field, S., Heatley, F. (2001). The cost-effectiveness of magnetic resonance imaging for investigation of the knee joint. *Health Technology Assessment*, 5(27), 1-95.
16. Patel, N., Bucknill, A., Ahearne, D., Denning, J., Desai, K., Watson, M. (2012). Early magnetic resonance imaging in acute knee injury: a cost analysis. *Knee Surgery,*

Sports Traumatology, Arthroscopy, 20, 1152-1158.

17. McDonald, I., Daly, J., Jelinek, V.M., Panetta, F., Gutman, J.M. (1996). Opening Pandora's box: the unpredictability of reassurance by a normal test result. *British Medical Journal*, 313(7053), 329-32.
18. Van Ravesteijn, H., van Dijk, I., Darmon, D., van de Laar, F., Lucassen, P., Hartman, T.O., van Weel, C., Speckens, A. (2012). The reassuring value of diagnostic tests: a systematic review. *Patient Education and Counseling*, 86(1), 3-8.
19. Verhagen, A., de Vet, H.C., de Bie, R.A., Kessels, A.G., Boers, M., Bouter, L.M., Knipschild, P.G. (1998). The Delphi list: a criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. *Journal of Clinical Epidemiology*, 51(12), 1235-41.
20. Verhagen, A., de Vet, H.C., de Bie, R.A., Boers, M., van den Brandt, P.A. (2001). The art of quality assessment of RCTs included in systematic reviews. *Journal of Clinical Epidemiology*, 54(7), 651-4.
21. Tulder, M.V., Suttorp, M., Morton, S., Bouter, L.M., Shekelle, P. (2009). Empirical evidence of an association between internal validity and effect size in randomized controlled trials of low-back pain. *Spine*, 34(16), 1685-92.
22. Higgins, J., Thompson, S.G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21, 1539-1558.
23. Cohen, J. *Satistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Earlbaum Associates, 1988.
24. Members, P.P. (2001). Philadelphia Panel Evidence-Based Clinical Practice Guidelines on Selected Rehabilitation Interventions: Overview and Methodology. *Physical Therapy*, 81(10), 1629-40.
25. *Cochrane Handbook for Systematic Reviews of Interventions*. The Cochrane Collaboration, 2011.
26. Guyatt, G., Oxman, A.D., Vist, G.E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., Schünneman, H.J. (2008). GRADE: an emerging consensus on ratin quality of evidence and strength recommendations. *British Medical Journal*, 336(7650):924-6.
27. Furlan, A., Pennick, V., Bombardier, C., Tulder, M.V., on behalf of the Editorial Board of the Cochrane Back Review Group. (2008). *Updated method guidelines for systematic reviews in the Cochrane Back Review Group*, C.B.R.G.E. Board Editor.
28. Ash, L., Modic, M.T., Obuchowski, N.A., Ross, J.S., Brant-Zawadzki, M.N., Grooff, P.N. (2008). Effects of diagnostic information, per se, on patient outcomes in acute radiculopathy and low back pain. *American Journal of Neuroradiology*, 29(6), 1098-103.
29. Cohen, S., Gupta, A., Strassels, S.A., Christo, P.J., Erdek, M.A., Griffith, S.R., Kurihara, C., Buckenmaier, C.C., Cornblath, D., Vu, T. (2012). Effect of MRI on Treatment Results or Decision Making in Patients With Lumbosacral Radiculopathy Referred for Epidural Steroid Injections. *Archive of Internal Medicine*, 172(2), 134-42.
30. Deyo, R., Diehl, A.K., Rosenthal, M. (1987). Reducing roentgenography use. Can patient expectations be altered? *Archive of Internal Medicine*, 147(1), 141-5.
31. Djais, N., Kalim, H. (2005). The role of lumbar spine radiography in the outcomes of patients with simple acute low back pain. *APLAR J Rheumatology*, 8, 45-50.
32. Gilbert, F., Grant, A.M., Gillan, M.G., Vale, L., Scott, N.W., Campbell, M.K., Wardlaw, D., Knight, D., McIntosh, E., Porter, R.W. (2004). Does early imaging influence management and improve outcome in patients with low back pain? A pragmatic randomised controlled trial. *Health Technology Assessment*, 8(17), 1-131.
33. Kendrick, D., Fielding, K., Bentley, E., Kerslake, R., Miller, P., Pringle, M. (2001). *Radiography of the lumbar spine in*

primary care patients with low back pain: randomised controlled trial. British Medical Journal, 322(7283), 400-5.

34. Kendrick, D., Fielding, K., Bentley, E., Miller, P., Kerslake R., Pringle, M. (2001). The role of radiography in primary care patients with low back pain of at least 6 weeks duration: a randomised (unblinded) controlled trial. *Health Technology Assessment*, 5(30), 1-69.
35. Kerry, S., Hilton, S., Dundas, D., Rink, E., Oakeshott, P. (2002). Radiography for low back pain: a randomised controlled trial and observational study in primary care. *British Journal of General Practice*, 52(479), 469-74.
36. Kerry, S., Hilton, S., Patel, S., Dundas, D., Rink, E., Lord, J. (2000). Routine referral for radiography of patients presenting with low back pain: is patients' outcome influenced by GPs' referral for plain radiography? *Health Technology Assessment*, 4(20), 1-119.
37. Bryan, S., Bungay, H.P., Weatherburn, G., Field, S. (2004). Magnetic resonance imaging for investigation of the knee joint: a clinical and economic evaluation. *International Journal of Technology Assessment in Health Care*, 20(2), 222-9.
38. Modic, M., Obuchowski, N.A., Ross, J.S., Brant-Zawadzki, M.N., Grooff, P.N., Mazanec, D.J., Benzel, E.C. (2005). Acute Low Back Pain and Radiculopathy: MR Imaging Findings anf Their Prognostic Role and Effect on Outcome. *Radiology*, 237, 597-604.
39. Garratt, A.M., Brealy, S., Gillespie, W.J., DAMASK (Direct Access to Magnetic Resonance Imaging: Assessment for Suspect Knees) Trial Team. (2004). Patient-assessed health instruments for the knee: a structured review. *Rheumatology (Oxford)*, 43(11), 1414-23.
40. Brealy, S., Atwell, C., Bryan, S., Coulton, S., Cox, H., Cross, B., Fylan, F., Garratt, A., Gilbert, F.J., Gillan, M.G.C., Hendry, M., Hood, K., Houston, H., King, D., Morton, V., Orchard, J., Robling, M., Russell, I.T., Torgerson, D., Wadsworth, V., Wilkinson, C. (2006). The DAMASK trial protocol: a pragmatic randomised trial to evaluate whether GPs should have direct access to MRI for patients with suspected internal derangement of the knee. *BMC Health Services Research*, 6, DOI: 10.1186/1472-6963-6-133.
41. Brealy, S., DAMASK (Direct Access to Magnetic Resonance Imaging: Assessment for Suspect Knees) Trial Team. (2007). Influence of magnetic resonance of the knee on GPs' decisions: a randomised trial. *British Journal of General Practice*, 57(541), 622-9.
42. Brealy, S., DAMASK (Direct Access to Magnetic Resonance Imaging: Assessment for Suspect Knees) Trial Team. (2008). Cost-effectiveness of magnetic resonance imaging of the knee for patients presenting in primary care. *Br J Gen Pract*, 58, e10-e16 DOI: 10.3399/bjgp08X342660.
43. Brealy, S., Andronis, L., Dale, V., Gibbon, A.J., Gilbert, F.J., Hendry, M., Hood, K., King, D., Wilkinson, C., DAMASK (Direct Access to Magnetic Resonance Imaging: Assessment for Suspect Knees) Trial Team. (2012). The effect of waiting times from general practitioner referral to MRI or orthopaedic consultation for the knee on patient-based outcomes. *Br J Radiol*, 85, e1134-9 DOI: 10.1259/bjr/12729937.
44. Laisné, F., Lecomte, C., Corbière, M. (2012). Biopsychosocial predictors of prognosis in musculoskeletal disorders: a systematic review of the literature. *Disability and Rehabilitation*, 34(5), 355-82.
45. Ramond, A., Bouton, C., Richard, I., Roquelaure, Y., Baufreton, C., Legrand, E., Huez, J.F. (2011). Psychosocial risk factors for chronic low back pain in primary care–a systematic review. *Family Practice*, 28(1), 12-21.

46. Pincus, T., Burton, A.K., Vogel, S., Field, A.P. (2002). A systematic review of psychological factors as predictors of chronicity/disability in prospective cohorts of low back pain. *Spine*, 27(5), E109-20.
47. Steenstra, I., Verbeek, J.H., Heymans, M.W., Bongers, P.M. (2005). Prognostic factors for duration of sick leave in patients sick listed with acute low back pain: a systematic review of the literature. *Occupational and Environmental Medicine*, 62(12), 851-60.
48. Jarvik, J., Maravilla, K.R., Haynor, D.R., Levitz, M., Deyo, R.A. (1997). Rapid MR imaging versus plain radiography in patients with low back pain: initial results of a randomized study. *Radiology*, 204, 447-54.
49. Schulz, K., Grimes, D.A. (2002). Blinding in randomised trials: hiding who got what. *Lancet*, 359, 696-700.
50. Kent, W. Randomisation and blinding in clinical trials. WebmedCentral CLINICAL TRIALS 2012; 3(5):WMC003289.
51. Malý, M., Vondra, V. (2006). Generic versus disease-specific instruments in quality-of-life assessment of chronic obstructive pulmonary disease. *Methods of Infromation in Medicine*, 45(2), 211-5.
52. Bombardier, C., Melfi, C.A., Paul, J., Green, R., Hawker, G., Wright, J., Coyte, P. (1995). Comparison of a generic and a disease-specific measure of pain and physical function after knee replacement surgery. *Medical Care*, 33(4), AS131-44.
53. Pengel, L.H., Herbert, R.D., Maher, C.G., et.al. (2003). Acute low back pain: systematic review of its prognosis. *British Medical Journal*, 327(7410), 323
54. Rinkus, K.M., Knaub, M.A. (2008). Clinical and Diagnostic Evaluation of Low Back Pain. *Seminars in Spine Surgery*, 20, 93-101
55. Wheeler, A.H. (1995). Diagnosis and management of low back pain and sciatica. *Am Fam Phys*, 52, 1333-41
56. Davis, P.C., Wippold, F.J., Cornelius, R.S., Angtuaco, E.J., Broderick, D.F., Brown, D.C., Garvin, C.F., Hartl, R., Holly, L., McConnell, C.T. Jr., Mechtler, L.L., Rosenow, J.M., Seidenwurm, D.J., Smirniotopoulos, J.G., Expert Panel on Neurologic Imaging. (2011). ACR Appropriateness Criteria® low back pain. [online publication]. Reston (VA): American College of Radiology (ACR), 8
57. Van der Waal, J.M., Bot, S.D., Terwee, C.B., Van der Windt, D.A., Scholten, R.J., Bouten, L.M., Dekker, J. (2005). Course and prognosis of knee copmlaints in general practice. *Arthritis and Rheumatism*, 53(6), 920-30
57. Belo, J.N., Berg, H.F., Klein Ikkink, A.J., Wildervanck-Dekker, C.M.J., Smorenburg, H.A.A.J., Draijer, L.W. (2010). Clinical guideline ‘traumatic knee complaints’ from the Dutch College of General Practitioners (Dutch). *Huisarts en Wetenschap*, 54, 147-158
58] Bauer, S.J., Hollander, J.E., Fuchs, S.H., Thode, H.C. Jr. (1995). A clinical decision rule in the evaluation of acute knee injuries. *Journal of Emerging Medicine*, 14(5), 611-615
59. Seaberg, D.C., Jackson, R. (1994). Clinical decision rule for knee radiographs. *American Journal of Emerging Medicine*, 12(5), 541-543
60. Stiell, I.G., Wells, G.A., Hoag, R.H., et.al. (1997). Implementation of the Ottawa Knee Rule for the use of radiography in acute knee injuries. *JAMA*, 278(23), 2075-2079
61. American College of Radiology ACR. Appropriatness Criteria; nontraumatic knee pain. *Radiology*, 2000; 215(suppl): 1-1511
62. Oei, E.H., Nikken, J.J., Verstijnen, A.C., Ginai, A.Z., Hunink, M.G. (2003). MR imaging of the menisci and cruciate ligaments: a systematic review. *Radiology*, 226(3), 837-848

63. Crawford, R., Walley, G., Bridgman, S., Maffuli, N. (2007). Magnetic resonance imaging versus arthroscopy in the diagnosis of knee pathology, concentrating on meniscal lesions and ACL tears: a systematic review. *British Medical Bulletin*, 84, 5-23
64. Mackenzie, R., Palmer, C.R., Lomas, D.J., Dixon, A.K. (2006). Magnetic resonance imaging of the knee: diagnostic performance studies. *Clinical Radiology,* 51(4), 251-257

Appendix 1. Search strategy

Pubmed and Medline Mesh terms

("Back Pain"[Mesh] OR Back Pain*[tiab] OR Back ache*[tiab] OR Backache*[tiab] OR "Rheumatic Disease"[Mesh] OR Rheumatic Disease*[tiab] OR Rheumat*[tiab] OR Enthesopath*[tiab] OR Osteoarthrit*[tiab] OR "Neck Pain"[Mesh] OR Neck Pain*[tiab] OR Neck ache*[tiab] OR Neckache*[tiab] OR Cervicalgia*[tiab] OR Cervical Pain*[tiab] OR "Shoulder pain"[Mesh] OR Shoulder pain*[tiab] OR "wrist injuries"[Mesh] OR wrist injur*[tiab] OR "hip injuries"[Mesh] OR Hip injury[tiab] OR "Patellofemoral Pain Syndrome"[Mesh] OR Patellofemoral Pain*[tiab] OR knee Pain*[tiab] OR "Knee injuries"[Mesh] OR Knee injur*[tiab] OR cruciate ligament tear*[tiab] OR Meniscus[tiab] OR "Foot injuries"[Mesh] OR Foot injur*[tiab] OR Achillis tendon[tiab] OR plantar fasciitis[tiab] OR musculoskeletal complaint*[tiab] OR musculoskeletal pain*[tiab] OR skeletal complaint*[tiab] OR skeletal pain*[tiab] OR muscular complaint*[tiab] OR muscular pain*[tiab] OR muscle complaint*[tiab] OR muscle pain*[tiab] OR muscles complaint*[tiab] OR muscles pain*[tiab])

AND

("Diagnostic Imaging"[mh] OR Diagnostic test*[tiab] OR mri[tiab] OR radiograph*[tiab] OR imaging*[tiab] OR Tomogra*[tiab] OR CT OR Ultrasonogra*[tiab] OR sonogra*[tiab])

AND

(randomized controlled trial[pt] OR controlled clinical trial[pt] OR randomized[tiab] OR placebo[tiab] OR randomly[tiab] OR trial[tiab] OR groups[tiab]) NOT (animals[mh] NOT humans[mh])

Embase Emtree terms

('musculoskeletal chest pain'/de OR 'musculoskeletal injury'/exp OR 'musculoskeletal pain'/de OR 'musculoskeletal disease'/exp OR 'musculoskeletal stiffness'/exp OR myalgia/exp OR 'wrist pain'/de OR 'wrist injury'/de OR 'elbow injury'/de OR 'shoulder pain'/de OR 'shoulder injury'/de OR 'neck pain'/de OR 'neck injury'/exp OR 'cervical spine injury'/de OR 'spine injury'/de OR 'low back pain'/de OR 'hip injury'/de OR 'hip pain'/de OR 'knee injury'/exp OR 'knee pain'/de OR 'ankle injury'/exp OR 'ankle pain'/de OR 'patellofemoral pain syndrome'/de OR osteoarthritis/exp OR 'foot injury'/exp OR 'plantar fasciitis'/de OR (((musculoskelet* OR skelet* OR muscular* OR muscle* OR wrist* OR elbow* OR shoulder* OR neck OR spine OR 'low back' OR 'lower back' OR lowback OR lumbar OR lumbal OR lumbosacral OR hip OR hips OR knee* OR ankle* OR loin OR cervical OR patellofemor* OR lumbosacroiliac OR ligament* OR foot OR feet) NEAR/3 (pain* OR ache* OR complaint* OR injur* OR syndrome* OR disorder* OR symptom* OR strain* OR rupture* OR menisc* OR achilles OR tendon* OR lesion* OR tear* OR failure*)) OR myalgia* OR neckache* OR cervicalgia* OR lumbago OR lumbagalg* OR lumbodynia* OR osteoarthr* OR arthritis OR arthrosis OR 'degenerative joint disease' OR 'osteo arthritis' OR 'plantar fasciitis'):ab,ti)

AND

('imaging and display'/de OR 'diagnostic imaging'/de OR thermography/de OR spectroscopy/exp OR scintigraphy/exp OR radiography/exp OR 'computer assisted tomography'/exp OR 'nuclear magnetic resonance imaging'/exp OR ultrasound/de OR echography/exp OR myelography/de OR thermography/de OR (imaging OR radioimaging OR thermogra* OR spectroscop* OR scintigra* OR laminoscintigra* OR scintillation* OR scintillogra* OR scintiphotogra* OR radiogra* OR electroradiogra* OR pneumoradiogra* OR radiophotogra* OR roentgen* OR rontgen* OR x-ray OR xray OR tomogram* OR ((cat OR ct) NEXT/1 scan*) OR mri OR nmri OR NMR OR ultraso* OR 'ultra sound' OR echogram* OR echoscop* OR echosound OR sonogram* OR ultrasonogram* OR myelogra* OR medullogra* OR thermogra* OR thermoscan* OR infrared OR 'ophthalmo diaphanoscopy' OR transillumination):ab,ti)

AND

((random* OR factorial* OR crossover* OR (cross NEXT/1 over*) OR placebo* OR ((doubl* OR singl*) NEXT/1 blind*) OR assign* OR allocat* OR volunteer*):ab,ti OR 'crossover procedure'/de OR 'double-blind procedure'/de OR 'randomized controlled trial'/de OR 'single-blind procedure'/de) NOT ([animals]/lim NOT [humans]/lim)

AND

('disease course'/exp OR 'therapy effect'/de OR 'pain assessment'/de OR reassurance/de OR 'daily life activity'/de OR 'ADL disability'/exp OR 'patient satisfaction'/de OR 'psychological aspect'/de OR anxiety/de OR 'cost effectiveness analysis'/de OR (convalescen* OR recover* OR deteriorate* OR 'disease course' OR prognis* OR relapse* OR (therap* NEAR/3 effect*) OR (pain NEAR/3 (assess* OR measure* OR score*)) OR reassur* OR (daily NEAR/3 (activit* OR function*)) OR ADL OR satisf* OR psycholog* OR anxiety*):ab,ti)

Appendix 2. PRISMA checklist

Section/topic	#	Checklist item	Reported on page #
Title			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
Abstract			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2
Introduction			
Rationale	3	Describe the rationale for the review in the context of what is already known.	3
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	3
Methods			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	NA
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	4, 16-17
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	4
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	4
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	4
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	4
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	4
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	4
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	4-5
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	4-5

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	5
Additional analysis	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	5
Results			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	6
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	6
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	App 3
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	7-11
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	7-11
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	7-11
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	7-11
Discussion			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	7-11
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	11-13
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	13
Funding			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	14

NA Not applicable, Fig Figure, App Appendix, p page