European Journal of Internal Medicine 26 (2015) 585–595

Contents lists available at ScienceDirect

European Journal of Internal Medicine

journal homepage: www.elsevier.com/locate/ejim

Review Article

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

Yasmaine H.J.M. Karel [a,b,*], Karin Verkerk [c], Silvio Endenburg [c], Sven Metselaar [c], Arianne P. Verhagen [b]

[a] Research Group Diagnostics, Avans University of Applied Sciences, Breda, The Netherlands
[b] Department of Family Medicine, Erasmus Medical Center, Rotterdam, The Netherlands
[c] Education for Physiotherapy, Rotterdam University of Applied Sciences, The Netherlands

ARTICLE INFO

*Article history:*
Received 6 March 2015
Received in revised form 15 June 2015
Accepted 27 June 2015
Available online 15 July 2015

*Keywords:*
Diagnostic tests
Musculoskeletal/connective tissue disorders
Back pain
Primary care
Radiology

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.



## 1. 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 a 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

*Abbreviations:* SD, standard deviation; CI, confidence interval; IQR, interquartile range; LBP, low back pain; Gen, generic; Spec, specific; QoL, quality of life; $I^2$, heterogeneity statistic; Df, degrees of freedom; 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; EQ-5D, EuroQol 5 dimensions; SIP, sickness impact profile; VAS, visual analog scale; ALBP, Aberdeen Low Back Pain Score; FABQ, Fear Avoidance Beliefs Questionnaire; HADS, Hospital Anxiety and Depression Scale; US, United States; UK, United Kingdom.

* Corresponding author at: Research Group Diagnostics, Avans University of Applied Sciences, Hogeschoollaan 1, 4818 CR Breda, The Netherlands. Tel.: +31 6 13533708.
*E-mail address:* yhjm.karel@avans.nl (Y.H.J.M. Karel).

http://dx.doi.org/10.1016/j.ejim.2015.06.018
0953-6205/

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.

## 2. Methods

### 2.1. 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.

### 2.2. 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.

### 2.3. 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].

### 2.4. 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.

### 2.5. 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 difference (SMD) was calculated and a risk ratio (RR) for dichotomous outcomes including 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 settings, 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.

### 2.6. 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.

## 3. Results

### 3.1. Results of the search and description of studies

Searching the databases resulted in 13,167 references (Fig. 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.