2

Systematic review of patient history and physical examination for diagnosing facet joint pain in patients with low back pain

Esther T Maas, Johan NS Juch, Raymond WJG Ostelo, J George Groeneweg,
Jan-Willem Kallewaard, Bart W Koes, Arianne P Verhagen,
Frank JPM Huygen, Maurits W van Tulder
Submitted

ABSTRACT

Introduction Patient history and physical examination are frequently used procedures to diagnose chronic low back pain (CLBP) originating from the facet joints, although the diagnostic accuracy is controversial. The aim of this systematic review is to determine the diagnostic accuracy of patient history and/or physical examination to identify CLBP originating from the facet joints using diagnostic blocks as reference standard.

Databases and data treatment We searched MEDLINE, EMBASE, CINAHL, Web of Science and the Cochrane Collaboration database from inception until February 2015. Two review authors independently selected studies for inclusion, extracted data and assessed risk of bias. We calculated sensitivity and specificity values, with 95% confidence intervals (95%CI).

Results Twelve studies were included, in which 129 combinations of index test and reference standard were presented. Most of these index tests have only been evaluated in single studies with a high risk of bias and therefore provide insufficient evidence about the diagnostic accuracy of these tests. Four studies evaluated the diagnostic accuracy of the combination of Revel's criteria. Because of the clinical heterogeneity (variation in the reference standard), results were not pooled. The published sensitivities ranged from 0.11 (95%CI 0.02-0.29) to 1.00 (95%CI 0.75-1.00), and the specificities ranged from 0.66 (95%CI 0.46-0.82) to 0.91 (95%CI 0.83-0.96).

Conclusions Due to clinical heterogeneity, no firm conclusions could be drawn about the diagnostic accuracy of patient history and/or physical examination compared to a diagnostic facet joint block.

INTRODUCTION

Chronic low back pain (CLBP) is a widespread problem with major social and economic consequences.[1,2] Over the last decades multiple structures in the lumbar spine have been considered a cause of CLBP. Goldthwait et al. first described the facet joints as the source of pain in 1911, after which Ghormley introduced the term 'facet syndrome' in 1933.[3,4] Facet joint pain is defined as pain resulting from any integral structure of the facet joints, including the fibrous capsule, synovial membrane, hyaline cartilage surfaces and bony articulation.[5] The prevalence of facet joint pain is estimated at 15% to 41% of the CLBP population, and varies widely in the literature depending on setting, definition and diagnostic procedures.[6-11]

History taking and physical examination are two commonly used procedures in diagnosing facet joint pain and are considered the index test in this review. The validity and reliability of physical examination in clinical practice have been a matter of controversy.[5,12] Physicians and therapists use the information gained from history taking and physical examination to decide on the use of further diagnostic tests. Imaging (plain radiography, computed tomographic or magnetic resonance imaging), and diagnostic blocks tests are tests subsequent to physical examination, and are used in clinical practice to diagnose, amongst others, facet joint pain. The evidence on the ability of radiologic imaging to predict response to diagnostic facet joint blocks and diagnose CLBP originating from the facet joints has been shown to be conflicting at best, and are therefore not used in this review.[13-24]

The most commonly used test to diagnose CLBP originating from the facet joints is the diagnostic block. The rationale for these blocks is to anesthetize a painful facet joint for the duration of the anaesthetic effect.[25,26] The diagnostic block is recommended in many guidelines, among others in the Dutch guideline for anaesthesiology.[24] Although not a gold standard, diagnostic blocks are currently the best tests available for diagnosing facet joint pan, and therefore, chosen as reference standard in this review.[26-28, 29,30]

The last systematic review, in which the evidence of diagnostic accuracy of patient history and/or physical examination was summarised, was published more than eight years ago. It is important to update this review, because including more recent publications may have an impact on the overall evidence. Furthermore, patient care can potentially be improved if patients' history and physical examination could be used to limit the need for further invasive diagnostic procedures and treatments. Also, during the last years the methods of

diagnostic systematic reviews evolved and this update will use current best methodology. This review will provide an overview of the current practice in diagnosing CLBP originating from the facet joints.

The objective of this review is to assess the diagnostic accuracy of patient history and physical examination compared to diagnostic blocks to identify chronic low back pain originating from the facet joints.

METHODS

Design

A systematic review of diagnostic accuracy studies.

Data sources and search strategy

We performed a database search using Medline (in OvidSP), EMBASE, CINAHL, Web of Science, Google Scholar and the Cochrane Collaboration database from their date of inception until February 11th 2015 using MeSH terms (Medline), thesaurus (EMBASE, CINAHL, Web of Science) and free-text words (Cochrane, Google Scholar). Search terms were related to the diagnostic accuracy, facet joint pain, index and reference tests. Complete search strategies for the databases are available on request of the author. No method filter was included, because using filters might lead to missing relevant studies.[31] From the results of the electronic search, the bibliographies of all systematic reviews and eligible diagnostic studies were reviewed. We contacted experts in the field of diagnostic testing in anaesthesiology and LBP to ensure that the search was comprehensive.

Study selection

Two review authors (JJ and EM) independently screened the search results based on title, key words and abstract. We obtained the full texts for hits that were considered relevant by at least one of the authors. Disagreements were resolved by consensus. In case of persistent disagreement or uncertainty, a third reviewer (AV) made the final decision. Reasons for exclusion were noted. For final inclusion the studies had to fulfil the following criteria:

- Study design: retrospective, prospective or cross-sectional studies.
- Data collection: studies designed using existing data as well as studies designed

specifically to address the diagnostic question with newly collected data were included. No restrictions with regard to year of publication or language were applied. Non-English language reports were categorized as 'awaiting assessment' until appropriate translation was obtained. Papers published only in abstract form, case reports, as well as animal and post-mortem studies were excluded.

- Study population: adult patients, of either sex, suffering from CLBP were included. Patients with acute trauma, fractures, malignancies, and inflammatory diseases were excluded.
- Index test: all history taking and physical examination tests in the included studies were analysed, in isolation and in combination.
- Reference standard: A diagnostic block of the medial branch of the dorsal ramus under X-ray or low volume intra-articular blocks, as recommended by the Dutch guideline for anaesthesiology.[24] There is some variation in the use of diagnostic blocks.[5,29] In general, pain reduction of 50% or more implies that the anesthetized joint is the source of the pain, but studies using other thresholds were also included. Single and double (to confirm the results of the first test) diagnostic blocks have comparable validity and were therefore both included as reference standard.[24,32]

Information on the inter- or intra-observer reproducibility of the tests, or both, if reported or referenced in the study were collected. The maximum time interval between the index test and reference standard was three months, as no change of symptoms in CLBP patients was expected in this time range.

Data extraction

Two reviewers (JJ and EM) developed and completed the data extraction form.

Data were extracted on:

- Author, date of publication, location, journal
- Study design
- Study population characteristics: basic demographics, number of patients (number eligible for the study, number enrolled in the study, number receiving index test and reference standard, number of whom the results are reported in the two-by-two table, reasons for withdrawal), inclusion and exclusion criteria, setting
- Index tests and reference standard characteristics: type of test, method of execution, cut-off-value (outcome scales), positivity thresholds
- Outcomes: true-positives, false-positives, true-negatives and false-negatives

Quality assessment

Two reviewers (JJ and EM) independently assessed the methodological quality using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool.[33,34] Signalling questions were used to assess the risk of bias in four domains: patient selection (three signalling questions), index tests (two signalling questions), reference standard (two signalling questions), and flow and timing (four signalling questions). We added one extra item to the QUADAS-2 tool to assess reliability: 'Were data on observer variation reported and within an acceptable range?'. When there was at least one 'no' or 'unclear' response to a signalling question for a given domain, we scored the risk of bias domain as high or unclear, respectively. Furthermore, concerns about applicability of the studies were scored on three domains: patient selection, index test, and reference standard.

Disagreements were resolved by discussion or by an independent third author (RO). No summary scores were used.[35,36] All domains covered by the QUADAS-2 tool were considered as potential covariates. That is, if an item was assumed to potentially influence the reported sensitivity and specificity, this item was incorporated in the bivariate analyses to examine the effect of this potential source of bias on the diagnostic accuracy of patient history and physical examination. The items of the QUADAS-2 tool are available on request of the author.

Data synthesis and analysis

Study-specific estimates of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), as well as sensitivity and specificity for each index test with 95% confidence intervals (CI) were calculated. In case of clinical homogeneity (same index tests and reference standards, and a comparable study population) a multivariate meta-analysis was conducted using a hierarchical summary receiver operating characteristic (HSROC) model. Pooled estimates of sensitivity and specificity and 95% CIs were calculated with this model.[37] The HSROC model is recommended for analysing sensitivity and specificity data reporting on more than one threshold for positive tests.[38] The HSROC model estimates the between-study variability (heterogeneity), allowing for the theoretically expected non-independence of sensitivity and specificity across studies, and can therefore be seen as an improvement of the simple summary Receiver Operating Characteristic (ROC) approach.[37,38]

For each index test, parameter estimates from the fitted model were inputted into the software program Revman5 to graph the ROC space. We investigated the potential

influence of differences in study population, index test and reference standard criteria, differences in time interval of the evaluation of pain reduction, and study design features (prospective versus retrospective).
Where studies were clinically too heterogeneous, no meta-analysis was performed, and only the range of sensitivity and specificity and the 95% CI (as published) were presented.

RESULTS

Study selection

The study flow diagram is presented in Figure 1. Database searching yielded 824 individual papers and four additional papers were identified through reference checking. After removal of duplicates and exclusions based on title and abstract, 26 full-text papers were assessed for eligibility. Fourteen papers were excluded for various reasons (reasons are shown in Figure 1), which resulted in 12 included papers. Two studies reported on the same study population,[6,39] and one study [40] was part of a larger diagnostic accuracy study.[41]

Description of the studies

The 12 included studies consisted of ten cross-sectional studies, one case-control study,42 and one retrospective cohort,[43] in which information was collected from medical records. The sample size of the studies ranged from 51 [44] to 259 [43] patients and the mean age ranged from 38[39] to 62[43] years of age.
A variety of patient history and physical examination items were used as index tests. Index tests found in more than one study were: non-centralization, onset of trauma, age>65 years, pain well relieved by recumbency, pain not exacerbated by coughing, pain not exacerbated by forward flexion, pain not exacerbated by extension, pain not exacerbated by rising from flexion, and pain not exacerbated with the extension-rotation test. Five studies reported findings on diagnostic accuracy on combinations of test results. Four studies reported on the Revel's criteria.[18,44-46] Index tests which were evaluated only in single studies include: a pain distribution pattern,[43] a clinical prediction rule,[41] the new lumbar facet sign[42] and many aspects of a physical examination. All index tests compared to the reference standards are available upon request of the author.
Seven studies used a single diagnostic block as reference standard. Three studies

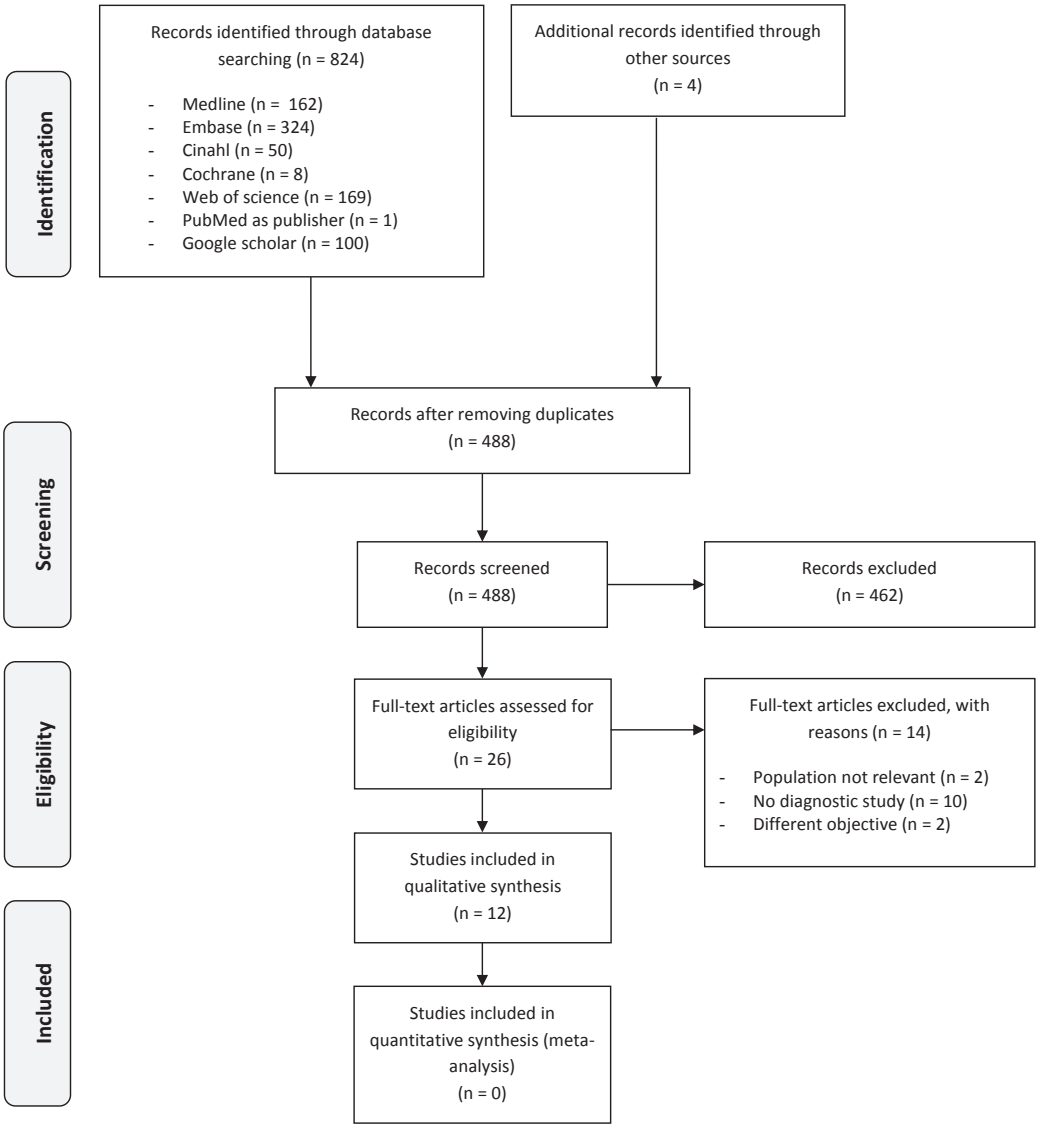

Figure 1. Study flow diagram

used a second confirmatory diagnostic block in patients with a positive first diagnostic block[11,42,46]. The cut-off point in which the diagnostic block was considered positive ranged from 'clinical improvement'[42] to 'more than 95% pain reduction 30 to 45 minutes after procedure'.[40]

Methodological quality

The results of the quality assessment are presented in Figure 2. The two reviewers (JJ and EM) agreed on 39 of the 48 risk of bias domains (80%) and on all applicability concern domains while scoring the QUADAS-2 tool.

Risk of bias: We scored patient selection 'unclear' or 'high risk of bias' in eight of the 12 studies. In four studies,[6,18,39,45] the patient selection was adequately described and scored 'low risk of bias'. Half of the studies scored 'low risk of bias' on description of the index test. Only the study of González et al[42] scored 'high risk of bias' on this domain as in this study the index test was interpreted with prior knowledge of the reference standard results. It remained unclear if a pre-specified threshold was used for the scoring of the index test in four studies, resulting in an 'unclear risk of bias' score. In eight studies it was unclear if the reference standard results were interpreted without knowledge of the index test results. We scored seven studies as 'high risk of bias' on the flow and timing domain. This was mostly due to the fact that not all patients received a reference standard or received the same reference standard.

Applicability concerns: For the assessment of applicability, there was no concern for nine studies that the included patients, setting, the conduct and interpretation of the index test, and the target condition (as defined by the reference standard) in each of the included studies did not match the review question. In one study[43] the patient selection was insufficiently described, in another study[41] the reference standard was not clearly described, and in a third study[11] the description of the index test was unclear.

Findings

The data extraction form of all 129 index tests and reference standard combination are available on request of the author. Ten index tests were evaluated in more than one study and are shown in Table 1. Because of clinical heterogeneity, a meta-analysis was not considered relevant.

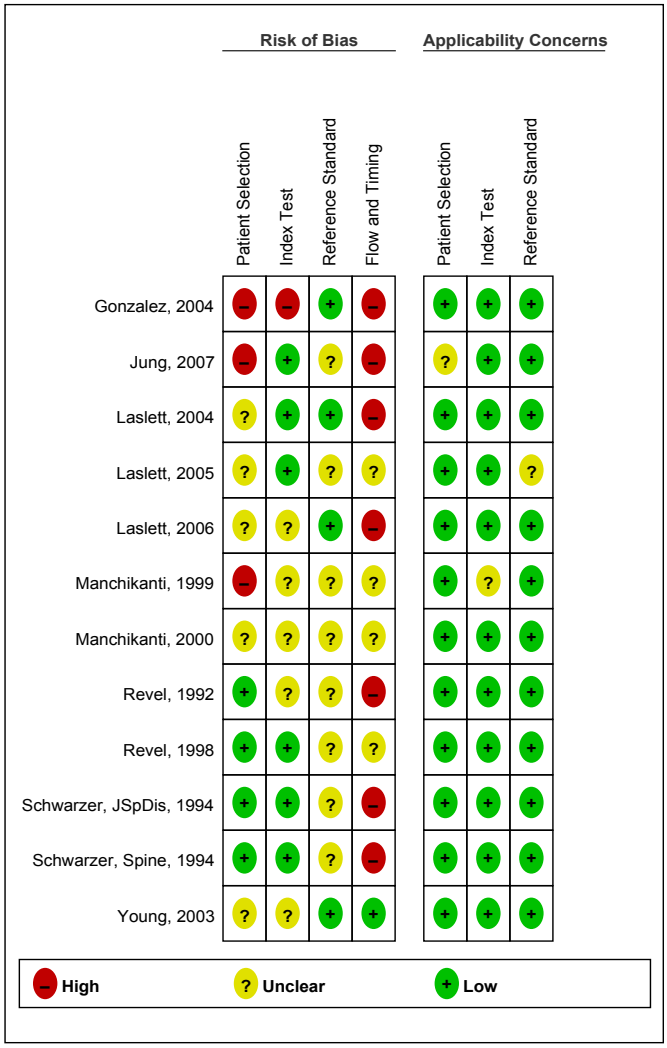

Figure 2. Quality assessment summary by the QUADAS-2 tool

Table 1. Diagnostic value of index tests for facet joint pain in two or more studies

Study	Reference standard	Sample size	TP	TN	FP	FN	Prior probability	Sensitivity (95%CI)	Specificity (95%CI)
	Non- centralization								
Laslett, 2006	≥95% pain reduction after diagnostic block	92	11	14	67	0	0.12	1.00 (0.74-1.00)	0.17 (0.11-0.27)
Young, 2003	≥80% pain reduction after diagnostic block	23	14	1	8	0	0.61	1.00 (0.78-1.00)	0.11 (0.02-0.44)
	Traumatic onset								
Manchikanti, 1999	≥75% pain reduction after double diagnostic block	120	29	31	35	25	0.45	0.54 (0.41-0.66)	0.47 (0.35-0.59)
Manchikanti, 2000	≥75% pain reduction after double diagnostic block	200	40	58	58	44	0.42	0.48 (0.37-0.58)	0.50 (0.41-0.59)
	Age >65 years								
Manchikanti, 1999	≥75% pain reduction after double diagnostic block	120	10	54	12	44	0.45	0.19 (0.10-0.31)	0.82 (0.71-0.89)
Manchikanti, 2000	≥75% pain reduction after double diagnostic block	200	18	99	17	66	0.42	0.21 (0.14-0.31)	0.85 (0.78-0.91)
Revel, 1998	≥75% pain reduction after lidocaine block	42	5	23	6	8	0.31	0.38 (0.18-0.64)	0.74 (0.62-0.90)
	Pain well relieved by recumbency								
Revel, 1992	≥75% pain reduction after lidocaine block	40	20	8	10	2	0.55	0.91 (0.72-0.97)	0.44 (0.25-0.66)
Revel, 1998	≥75% pain reduction after lidocaine block	42	12	7	22	1	0.31	0.92 (0.67-0.99)	0.24 (0.12-0.42)
	Pain not exacerbated by coughing								
Manchikanti, 2000	≥75% pain reduction after double diagnostic block	200	76	15	101	8	0.42	0.90 (0.82-0.95)	0.13 (0.08-0.20)
Revel, 1992	≥75% pain reduction after lidocaine block	40	18	9	9	4	0.55	0.82 (0.62-0.93)	0.50 (0.29-0.71)
Revel, 1998	≥75 pain reduction after lidocaine block	42	13	10	19	0	0.31	1.00 (0.77-1.00)	0.35 (0.20-0.53)
	Pain not exacerbated by forward flexion								
Manchikanti, 2000	≥75% pain reduction after double diagnostic block	200	13	95	21	71	0.42	0.15 (0.09-0.25)	0.82 (0.74-0.88)
Revel, 1992	≥75% pain reduction after lidocaine block	40	14	14	4	8	0.55	0.64 (0.41-0.83)	0.78 (0.52-0.94)
Revel, 1998	≥75% pain reduction after lidocaine block	42	13	14	15	0	0.31	1.00 (0.77-1.00)	0.48 (0.31-0.66)
	Pain not exacerbated by extension								
Manchikanti, 2000	≥75% pain reduction after double diagnostic block	200	8	100	16	76	0.42	0.10 (0.05-0.18)	0.86 (0.79-0.91)
Revel, 1992	≥75% pain reduction after lidocaine block	40	12	13	5	10	0.55	0.55 (0.35-0.73)	0.72 (0.49-0.88)
Revel, 1998	≥75% pain reduction after lidocaine block	42	12	18	11	1	0.31	0.92 (0.67-0.98)	0.62 (0.44-0.77)

Study	Reference standard	Sample size	TP	TN	FP	FN	Prior probability	Sensitivity (95%CI)	Specificity (95%CI)
	Pain not exacerbated by rising from flexion								
Manchikanti, 2000	≥75% pain reduction after double diagnostic block	210	48	59	63	40	0.42	0.55 (0.44-0.65)	0.48 (0.39-0.57)
Revel, 1992	**≥75% pain reduction after lidocaine block**	40	17	10	8	5	0.55	0.77 (0.57-0.90)	0.56 (0.34-0.75)
Revel, 1998	≥75% pain reduction after lidocaine block	42	13	17	12	0	0.31	1.00 (0.77-1.00)	0.59 (0.41-0.74)
	Pain not exacerbated with the extension-rotation test								
Manchikanti, 2000	**≥75% pain reduction after double diagnostic block**	200	57	35	81	27	0.42	0.68 (0.57-0.77)	0.30 (0.23-0.39)
Revel, 1998	≥75% pain reduction after lidocaine block	42	10	14	15	3	0.31	0.76 (0.50-0.92)	0.48 (0.31-0.66)
	Revel's criteria combined*	200	40	58	58	44	0.42	0.48 (0.37-0.58)	0.50 (0.41-0.59)
Laslett, 2004	75% pain reduction after diagnostic block	116	3	81	8	24	0.23	0.11 (0.02-0.29)	0.91 (0.83-0.96)
Manchikanti, 2000	≥75% pain reduction after double diagnostic block	40	14	16	2	8	0.55	0.13 (0.07-0.22)	0.84 (0.76-0.90)
Revel, 1992	≥75% pain reduction after Lidocaine block	40	14	16	2	8	0.55	0.64 (0.41-0.83)	0.89 (0.65-0.99)
Revel, 1998	≥75% pain reduction after Lidocaine block	42	13	19	10	0	0.31	1.00 (0.75-1.00)	0.66 (0.46-0.82)

Abbreviations: TP= true positives, FP= false positives, TN= true negatives, FN= false negatives
** Revel's criteria include (1) age over 65 years, (2) pain well relieved by recumbency and pain not exacerbated by (3) coughing, (4) forward flexion, (5) extension, (6) rising from flexion, (7) extension-rotation. Revel's criteria are positive with five or more clinical characteristics*

Diagnostic value of index tests for facet joint pain described in two or more studies

Non-centralization

The non-centralization phenomenon [47,48] was evaluated as index test in two studies.[40,49] Both studies used a single diagnostic block as reference standard, while Laslett et al. used a cut-off point of 80% pain reduction and Young et al. of 95% pain reduction.[40,49] Both studies did not have false negative cases, which resulted in a sensitivity of 1.00. Specificity was poor in both studies: 0.17 (95%CI 0.11-0.27) in Laslett et al.;[40] 0.11 (95% CI 0.02-0.44) in Young et al.[49]

Traumatic onset

Trauma as a cause for facet joint pain was evaluated in two studies.[11,46] The index test was compared to a controlled double diagnostic block in both studies. Sensitivity and specificity were poor with a maximum sensitivity of 0.54 (95%CI 0.37- 0.58) and maximum specificity of 0.50 (95%CI 0.41-0.59).

Revel's criteria separately

Revel's criteria include (1) age over 65 years, (2) pain well relieved by recumbency and (3) pain not exacerbated by coughing, (4) forward flexion, (5) extension, (6) rising from flexion, (7) extension-rotation.[45] All tests were studied by Revel[18,45] and/or Manchikanti.[11,46] Because of the heterogeneity in the reference standard (the use of single and double diagnostic blocks), no pooled results are presented. Sensitivity ranged from 0.15 (95%CI 0.09-0.25) to 1.00 (95%CI 0.77-1.00); specificity ranged from 0.13 (95%CI 0.08-0.20) to 0.86 (95%CI 0.79-0.91).

Revel's criteria combined

Four studies evaluated the performance of the combined Revel's criteria (positive with five or more clinical characteristics).[18,44-46] Because of the clinical heterogeneity (variation in the reference standard), results were not pooled. Sensitivity ranges from 0.11 (95%CI 0.02-0.29) to 1.00 (95%CI 0.75-1.00), and the specificity ranges from 0.66 (95%CI 0.46-0.82) to 0.91 (95%CI 0.83-0.96) (Figure 3).

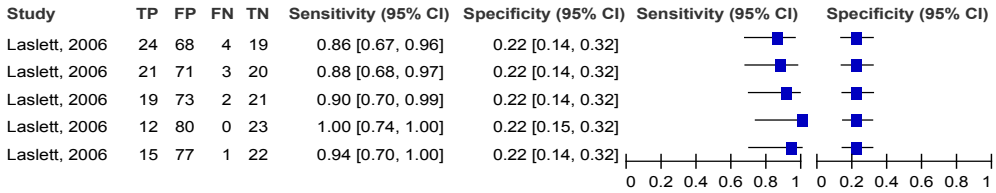

Figure 3. Revel's criteria – reference test (single diagnostic block in the studies of Laslett, Revel (1992) and Revel (1998). Double diagnostic block in the study of Manchikanti)

Diagnostic values of index tests for facet joint pain in one study

For 48 of the 129 index test and reference standard combinations that were described in only one study, it was impossible to construct a 2x2 table.

In the study of Laslett et al,[40] the extension-rotation test was compared to five thresholds of reference standards: 75%, 80%, 85%, 90% and 95% pain reduction after a diagnostic block. Almost no change in specificity was shown due to the very little change in true negatives. Neither was a difference in sensitivity shown (See Figure 4).

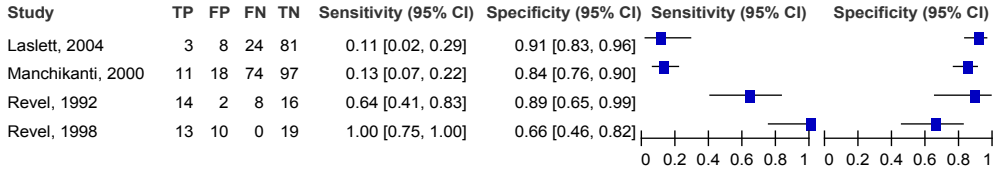

Figure 4. Extension-rotation test – reference test: 1. 75%, 2. 80%, 3. 85%, 4. 90% and 5. 95% pain reduction after a diagnostic block (Study by Laslett et al, 2006 (N=120))

DISCUSSION

This review included 12 diagnostics accuracy studies, including 129 combinations of patient history and/or physical examination tests with diagnostics blocks as reference standard, to identify the facet joints as the source of CLBP. Most index tests have only been evaluated in small single studies with a high risk of bias and therefore provide insufficient evidence about the diagnostic accuracy of these tests. The diagnostic accuracy of Revel's criteria was evaluated in four studies, which only concerned two research teams. Sensitivity ranged from 0.15 (95%CI 0.09-0.25) to 1.00 (95%CI 0.77-1.00); specificity ranged from 0.13 (95%CI 0.08-0.20) to 0.86 (95%CI 0.79 -0.91). No meta-analyses were performed due to clinical heterogeneity.

Factors affecting interpretation

One of the reasons for clinical heterogeneity in this review could be the wide range in prior probabilities of having CLBP originating from the facet joints in the included studies (11.9% to 55.0%).[18,40] Sensitivity and specificity can vary between studies with different prior probabilities of disease.[50] Furthermore, setting and patient selection was only described clearly in four of the 12 included studies. This could have affected the results as well, as patients in the other eight included studies may not have been a representative sample. Moreover, it is important to state that the results of this review can only be interpreted in a secondary care context. Patient history and physical examination are conducted in multiple settings, but diagnostic blocks are used as predictor for the response on RF denervation at the pain clinic which is positioned in a secondary care setting.

Secondly, there is no gold standard for diagnosing facet joint pain. In this review the anaesthetic diagnostic block was used as reference test. However, little evidence is available that truly justifies the use of diagnostic blocks as the gold standard but merely denotes that it is currently the best tool available. Still it is important to stress that the lack of a gold standard represents a flaw in diagnostic studies. As the index test can never perform better than the reference standard, its value may be underestimated.[51] Despite the lack of validity and the chance of false-positive tests,[28,29] diagnostic blocks are currently the best tests available for diagnosing CLBP originating from the facet joints.[24]

Thirdly, recent studies included in this review showed large variation in type of index tests. Very few studies were reproduced to confirm results. This resulted in many index tests that were only evaluated in single studies. The results of this review should therefore be interpreted with caution. Additionally, the reliability of the index tests and the (inter- and intra-) observer variation may have affected the results due to lack of consensus in procedures and thresholds. Risk of bias was considered high on the item 'index test' in six studies because of a lack of definition of a positive or negative test result.

Reliability

This review focused on the diagnostic performance (i.e. validity) of patient history and/or physical examination in patients with CLBP. The reliability of these tests was outside the scope of this review. However, adequate reliability (inter- and intra-observer agreement) is a prerequisite for good performance of diagnostic tests.

Only three studies provided some information on reliability.[6,39,41] Laslett et al (2005) reported weak inter-observer agreement for the physiotherapist diagnosis of facet joint pain with a kappa value of 0.31 (95%CI 0.18-0.44). Both studies of Schwarzer et al (1994) reported substantial to excellent inter-observer agreement in physical examination between the principal investigator and the other investigators. The κ scores for comparisons between observers ranged between 0.69 and 1.00.

Strengths and limitations

A strength of this review is the assessment of methodological quality by two independent assessors using the QUADAS-2 tool. Careful assessment of the quality of included studies is essential in a systematic review and increases the validity of the review. Secondly, we used a comprehensive search strategy. The use of search filters was avoided; which minimized the risk of missing relevant studies.

The most important limitation of this review was the relatively small number of included studies. Since the last systematic review on this topic in 2007,[12] six new studies have been published and were included in this review. This doubled the number of studies regarding facet joint pain. However, clinical heterogeneity among the included studies was substantial. Consequently, the improvement of evidence on the diagnostic accuracy of patient history and physical activity in diagnosing facet joint pain compared to the previous review in 2007 is limited.

A second limitation of this review is the poor description of the performed physical examination in most studies. It often remained unclear what thresholds were used to define positive test results. This may have led to different positive and negative test results between studies.

Implications for practice

The importance of the findings should be interpreted in light of its clinical consequences. To cite Revel et al. 1998[45]: '*A set of five clinical characteristics could be used in future randomized controlled studies as a criterion for selecting patients with probable painful facet joints that will be well relieved by facet joint anaesthesia. However, these clinical characteristics should not be considered as diagnostic criteria for low back pain of facet joint origin in clinical practice with individual patients*'. According to the results in the updated review, not much has changed in the past 17 years and the association between performing physical examination and a positive diagnostic block to identify

facet joint pain remains inconclusive. The diagnostic accuracy of patient history and/or physical examination for identifying other sources of back pain, such as disc herniation or sacroiliac joint pain is also poor, based on other studies.[12,52] The diagnostic value of physical examination tests in primary care populations and other general, unselected patient groups is still unclear as well, as evidence from these settings is scarce.

Implications for research

New studies on this topic have not led to more clarity, only to more heterogeneity. The evidence for the diagnostic accuracy of patient history and/or physical examination identified by this review is still inconclusive. To provide more profound evidence on the role of patient history and physical examination, there is a strong need for good quality and accurately reported prospective cohort studies. These studies should especially focus on investigating diagnostic accuracy of existing, commonly used tests, rather that introducing new tests. Implementation of the STARD guidelines will improve reporting of diagnostic studies in future research.[53,54]

Conclusions

Studies comparing Revel's criteria to a diagnostic facet joint block found conflicting evidence. Other patient history and/or physical examination items have mostly been investigated only in single studies with high risk of bias. Therefore, there is insufficient evidence about the diagnostic accuracy of these tests. Based on this review, no firm conclusions can be drawn about the association between patient history and/or physical examination and diagnostic facet joint blocks. Further high quality, more accurately reported diagnostic studies are required to confirm results in single studies.

Acknowledgements

We thank Wichor Bramer for his useful help in designing the search strategy.

REFERENCES

1. Martin BI, Deyo RA, Mirza SK, et al. Expenditures and health status among adults with back and neck problems. *JAMA.* 2008;299:656-664.
2. Manchikanti L, Singh V, Datta S, Cohen S, Hirsch J. Comprehensive review of epidemiology, scope, and impact of spinal pain. *Pain Physician*. 2009;12:35-70.
3. Goldthwait J. The Lumbo-Sacral Articulation; An Explanation of Many Cases of" Lumbago,"" Sciatica" and Paraplegia. *Boston Med Surg J*. 1911;164:365-372.
4. Ghormley R. Low back pain with special reference to the articular facets, with presentation of an operative procedure. *JAMA*. 1933;101:1773-1777.
5. Cohen S, Raja S. Pathogenesis, diagnosis, and treatment of lumbar zygapophysial (facet) joint pain. *Anesthesiology*. 2007;106:591-614.
6. Schwarzer A, Aprill C, Derby R, Fortin J, Kine G, Bogduk N. Clinical features of patients with pain stemming from the lumbar zygapophysial joints. *Spine*. 1994;19:1132-1137.
7. Manchikanti L, Boswell M, Singh V, Pampati V, Damron K, Beyer C. Prevalence of facet joint pain in chronic spinal pain of cervical, thoracic, and lumbar regions. *BMC Musculoskelet Disord*. 2004;5:15.
8. DePalma M, Ketchum J, Saullo T. What Is the Source of Chronic Low Back Pain and Does Age Play a Role? *Pain Med*. 2011;12:224-233.
9. Hicks G, Morone N, Weiner D. Degenerative lumbar disc and facet disease in older adults: prevalence and clinical correlates. *Spine*. 2009;34:1301-1306.
10. Eubanks J, Lee M, Cassinelli E, Ahn N. Prevalence of lumbar facet arthrosis and its relationship to age, sex, and race: an anatomic study of cadaveric specimens. *Spine*. 2007;32:2058-2062.
11. Manchikanti L, Pampati V, Fellows B, Bakhit C. Prevalence of lumbar facet joint pain in chronic low back pain. *Pain Physician*. 1999;2:59-64.
12. Hancock M, Maher C, Latimer J, et al. Systematic review of tests to identify the disc, SIJ or facet joint as the source of low back pain. *Eur Spine J*. 2007;16:1539-1550.
13. Lewinnek G, Warfield C. Facet joint degeneration as a cause of low back pain. *Clin Orthop Relat R*. 1986;213:216-222.
14. Carrera G. Lumbar facet joint injection in low back pain and sciatica: preliminary results. *Radiology.* 1980;137:665-667.
15. Helbig T, Lee C. The lumbar facet syndrome. *Spine*. 1988;13:61.
16. Carrera G, Williams A. Current concepts in evaluation of the lumbar facet joints. *Crit Rev Diagn Imag*. 1984;21:85-104.
17. Dolan A, Ryan P, Arden N, et al. The value of SPECT scans in identifying back pain likely to benefit from facet joint injection. *Rheumatology*. 1996;35:1269-1273.
18. Revel M, Listrat V, Chevalier X, et al. Facet joint block for low back pain: identifying predictors of a good response. *Arch Phys Med Rehab*. 1992;73:824-828.
19. Jackson R, Jacobs R, Montesano P. Facet joint injection in low-back pain. A prospective statistical study. *Spine*. 1988;13:966-971.
20. Murtagh F. Computed tomography and fluoroscopy guided anesthesia and steroid injection in facet syndrome. *Spine*. 1988;13:686-689.
21. Fairbank J, Park W, McCall I, O'Brien J. Apophyseal injection of local anesthetic as a diagnostic aid in primary low-back pain syndromes. *Spine*. 1981;6:598-605.

22. Raymond J, Dumas J, Lisbona R. Nuclear imaging as a screening test for patients referred for intraarticular facet block. *J Can Assoc Radiol*. 1984;35:291-292.
23. Schwarzer A, Wang S, O'Driscoll D, Harrington T, Bogduk N, Laurent R. The ability of computed tomography to identify a painful zygapophysial joint in patients with chronic low back pain. *Spine*. 1995;20:907-912.
24. Itz C, Willems P, Zeilstra D, Huygen F. Dutch Multidisciplinary Guideline for Invasive Treatment of Pain Syndromes of the Lumbosacral Spine. *Pain Practice*. 2016; 16(1):90-110.
25. Bogduk N. The innervation of the lumbar spine. *Spine*. 1983;8(3):286-293.
26. Manchikanti L, Pampati V, Fellows B, Bakhit C. The diagnostic validity and therapeutic value of lumbar facet joint nerve blocks with or without adjuvant agents. *Curr Pain Headache R*. 2000;4:337-344.
27. Dreyfuss P, Dreyer S, Vaccaro A. Lumbar zygapophysial (facet) joint injections. *Spine J*. 2003;3:50-59.
28. Schwarzer A, Aprill C, Derby R, Fortin J, Kine G, Bogduk N. The false-positive rate of uncontrolled diagnostic blocks of the lumbar zygapophysial joints. *Pain*. 1994;58:195-200.
29. Cohen S, Huang J, Brummett C. Facet joint pain—advances in patient selection and treatment. *Nat R Rheumatol*. 2012;9:101-116.
30. Hogan Q, Abram S. Neural blockade for diagnosis and prognosis: a review. Anesthesiology. 1997;86:216-241.
31. Beynon R, Leeflang MM, McDonald S, et al. Search strategies to identify diagnostic accuracy studies in MEDLINE and EMBASE. *The Cochrane Library*. 2013.
32. van Wijk R, Geurts J, Wynne H, et al. Radiofrequency denervation of lumbar facet joints in the treatment of chronic low back pain: a randomized, double-blind, sham lesion-controlled trial. *Clin J Pain*. 2005;21:335-344.
33. Whiting P, Rutjes A, Reitsma J, Bossuyt P, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol*. 2003;3(1):25-37.
34. Whiting P, Rutjes A, Westwood M, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155:529-536.
35. Whiting P, Harbord R, Kleijnen J. No role for quality scores in systematic reviews of diagnostic accuracy studies. *BMC Med Res methodol*. 2005;5(1):19.
36. Jüni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA*. 1999;282:1054-1060.
37. Reitsma J, Glas A, Rutjes A, Scholten R, Bossuyt P, Zwinderman A. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol*. 2005;58:982-990.
38. Dukic V, Gatsonis C. Meta-analysis of Diagnostic Test Accuracy Assessment Studies with Varying Number of Thresholds. *Biometrics*. 2003;59(4):936-946.
39. Schwarzer A, Derby R, Aprill C, Fortin J, Kine G, Bogduk N. Pain from the lumbar zygapophysial joints: a test of two models. *J Spinal Disord Tech*. 1994;7:331-336.
40. Laslett M, McDonald B, Aprill C, Tropp H, Őberg B. Clinical predictors of screening lumbar zygapophyseal joint blocks: development of clinical prediction rules. *Spine J*. 2006;6:370-379.
41. Laslett M, McDonald B, Tropp H, Aprill C, Öberg B. Agreement between diagnoses reached by clinical examination and available reference standards: a prospective study of 216 patients with lumbopelvic pain. *BMC Musculoskelet Disord*. 2005;6:28.

42. González J. Síndrome facetario lumbar. Nuevo signo de diagnóstico clínico. *Rehabilitación*. 2004;38:168-174.
43. Jung J-H, Kim H-I, Shin D-A, et al. Usefulness of pain distribution pattern assessment in decision-making for the patients with lumbar zygapophyseal and sacroiliac joint arthropathy. *J Korean Med Sci*. 2007;22:1048-1054.
44. Laslett M, Öberg B, Aprill C, McDonald B. Zygapophysial joint blocks in chronic low back pain: a test of Revel's model as a screening test. *BMC Musculoskeletal Disord*. 2004;5:43.
45. Revel M, Poiraudeau S, Auleley G, et al. Capacity of the clinical picture to characterize low back pain relieved by facet joint anesthesia: Proposed criteria to identify patients with painful facet joints. *Spine*. 1998;23:1972-1976.
46. Manchikanti L, Pampati V, Fellows B, Baha A. The inability of the clinical picture to characterize pain from facet joints. *Pain Physician*. 2000;3:158-166.
47. Sufka A, Hauger B, Trenary M, et al. Centralization of low back pain and perceived functional outcome. *J Orthop Sport Phys*. 1998;27:205-212.
48. Werneke M, Hart D. Centralization phenomenon as a prognostic factor for chronic low back pain and disability. *Spine*. 2001;26:758-764.
49. Young S, Aprill C, Laslett M. Correlation of clinical examination characteristics with three sources of chronic low back pain. *Spine J*. 2003;3:460-465.
50. Leeflang MM, Rutjes AW, Reitsma JB, Hooft L, Bossuyt PM. Variation of a test's sensitivity and specificity with disease prevalence. *CMAJ*. 2013;185(11):E537-E544.
51. Riegelman RK. Studying a study and testing a test: how to read the medical evidence. Lippincott Williams & Wilkins; 2005.
52. Van der Windt D, Simons E, Riphagen I, et al. Physical examination for lumbar radiculopathy due to disc herniation in patients with low-back pain. *Cochrane Database Syst Rev*. 2010:Feb 17;12:CD007431.
53. Smidt N, Rutjes A, Van der Windt D, et al. The quality of diagnostic accuracy studies since the STARD statement Has it improved? *Neurology*. 2006;67:792-797.
54. Bossuyt P, Reitsma J, Bruns D, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Ann Int Med*. 2003;138:1-12.