



Bachelorthesis zur Erreichung des FH-Diploms als Bachelor of Science HES-SO
in Physiotherapie

HES-SO Valais Wallis Bereich Gesundheit & Soziale Arbeit

**Reliabilität und Validität von
Schmerzprovokationstests des Iliosakralgelenks:
eine systematische Review**

Erarbeitet von: Christine Bittel und Leonie von Arx

Klasse: BA07

Unter Betreuung von: Anne-Gabrielle Mittaz Hager

Leukerbad, 15. Juli 2010

Zusammenfassung

Hintergrund

Das Iliosakralgelenk ist trotz seiner geringen Mobilität häufig Ursache von Beschwerden in der Rücken-, Becken- und Beinregion. Die Schwierigkeit besteht darin, eine valide und reliable Diagnose zu stellen, welche eine Unterscheidung einer ISG- Problematik von anderen Problematiken mit ähnlichen Symptomen erlaubt. Da bis heute kein Goldstandard für die ISG-Diagnostik existiert, ist die Medizin auf reliable und valide Mobilitäts- und Schmerzprovokationstests angewiesen.

Ziele

Das Ziel unserer Review ist es, einen Überblick über die Reliabilität und Validität von verschiedenen Testkombinationen von Schmerzprovokationstests zu geben.

Methoden

Die systematische Literatursuche wurde in den Datenbanken Cochrane, Pubmed, Cinahl, Psychinfo und Web of Science durchgeführt. Das Auswahlverfahren der Artikel, sowie die Bewertung der methodologischen Qualität wurden mittels des Quadas-Tools von den zwei Review-Verfasserinnen unabhängig durchgeführt.

Resultate

Aus den sieben eingeschlossenen Studien kann gesamthaft keine Testkombination den anderen übergeordnet werden. Die Testkombination von Laslett et al. 2003 weist gleichzeitig eine hohe Sensitivität (0.91) und Spezifität (0.87) für drei oder mehr positive Tests auf und eignet sich daher am besten für eine valide diagnostische Entscheidung. Diese Testkombination besteht aus folgenden Tests: Distractionstest, Thigh thrust Test, Gaenslen Test, Kompressionstest und Sacral thrust Test.

Diskussion

Aufgrund der mangelnden Vergleichbarkeit der Resultate sowie der Testdurchführung ist eine Beurteilung der Reliabilität und der Validität der Testkombinationen nur begrenzt möglich. Weitere Studien mit besserer Vergleichbarkeit müssen durchgeführt werden.

Schlussfolgerung

Obschon die einzelnen Testkombinationen zum Teil eine bessere Reliabilität und Validität als Einzeltests aufweisen, ist ihre Aussagekraft aufgrund mangelnder Forschungsarbeit zu niedrig, um einen Referenztest zu ersetzen.

Schlüsselwörter: Reliabilität, Validität, Schmerzprovokationstests, Iliosakralgelenk

Résumé

Introduction

Malgré sa faible mobilité physiologique, l'articulation sacro-iliaque est souvent à l'origine de plaintes dans la région du dos, du bassin et des membres inférieures. Le diagnostic valide et fiable qui permet de distinguer une problématique sacro-iliaque d'une autre pathologie représente la principale difficulté. A cause de l'absence de «gold standard» pour le diagnostic de l'articulation sacro-iliaque, la médecine est dépendante des tests de mobilité et de provocation de la douleur, dont la fiabilité et la validité sont discutés.

Objectif

L'objectif de notre revue est de donner une vue globale sur la fiabilité et la validité de différentes combinaisons de tests de provocation de la douleur.

Méthode

Notre recherche systématique de la littérature a été effectuée dans les bases de données Cochrane, Pubmed, Cinahl, Psychinfo et Web of Science. Le procédé de sélection des articles ainsi que l'évaluation de la qualité méthodologique, évaluée à l'aide du Quadas-Tool, ont été faits de manière indépendante.

Résultats

Parmi les sept études incluses, aucune combinaison de tests n'est globalement meilleure qu'une autre. La combinaison de test dans l'étude de Laslett et al. (2003) montre une haute sensibilité (0.91) ainsi qu'une haute spécificité (0.87) si trois tests ou plus sont positifs et convient donc pour prendre une décision diagnostique valide. Cette combinaison de test est constituée des tests suivants: test de distraction, Thigh thrust Test, test de Gaenslen, test de compression et Sacral thrust Test.

Discussion

Une évaluation de la fiabilité et de la validité des combinaisons de tests n'est possible que partiellement à cause du manque de comparabilité des résultats. Il est nécessaire de faire des études ultérieures avec une meilleure comparabilité des résultats.

Conclusion

Malgré les combinaisons de tests qui montrent une meilleure fiabilité et validité que les tests uniques, le manque de recherches ne leur permet pas d'avoir une force d'expression suffisante pour remplacer un test de référence.

Mots-clés: fiabilité, validité, tests de provocation de la douleur, articulation sacro-iliaque

Danksagung

Wir danken allen, die uns bei der Realisierung unserer Bachelorarbeit in irgendeiner Form unterstützt haben. Trotzdem möchten wir es nicht unterlassen, einige Personen speziell zu erwähnen.

Ein herzliches Dankeschön an all die Lehrer, die uns während der letzten zwei Jahre auf unsere Bachelorarbeit vorbereitet und uns die nötigen Grundkenntnisse vermittelt haben.

Ohne eine Grossauswahl an Literatur wäre die Realisierung unserer Arbeit nicht möglich gewesen. Deshalb gebührt ein besonderer Dank den Autoren, die uns Literatur zur Verfügung gestellt haben, insbesondere Mark Laslett.

Ein weiterer Dank geht an unsere Korrekturleser Albin Bittel, Stephanie Hofmann, Denise Jeitziner und Robert von Arx, die uns geholfen haben, unserer Arbeit den letzten Schliff zu geben.

Ein grosses Dankeschön richten wir an unsere Betreuerin Anne-Gabrielle Mittaz Hager, die uns während des letzten Jahres tatkräftig zur Seite stand. Als wir manchmal ratlos waren, sei es bei der Themenwahl ganz zu Beginn oder bei allfälligen Fragen während des Schlusspurts, konnten wir stets auf ihre Unterstützung zählen und von ihrer persönlichen Erfahrung bezüglich unseres Themas profitieren. Dies wissen wir sehr zu schätzen. Danke Gaby!

Inhaltsverzeichnis

1.	Einleitung.....	1
1.1.	Theoretischer Hintergrund	2
1.1.1.	Deskriptive und funktionelle Anatomie	2
1.1.2.	Ätiologie des ISG-Syndroms	3
1.1.2.1.	Funktionelle Störungen	3
1.1.2.2.	Degenerative Veränderungen	4
1.1.2.3.	Entzündliche Veränderungen	4
1.1.3.	Diagnostik des ISG	4
1.1.3.1.	Schmerzprovokationstests des ISG	5
1.1.3.2.	Invasive und bildgebende Verfahren des ISG	5
1.1.4.	Schmerz.....	6
1.1.5.	Outcomes	7
1.2.	Ziel	8
2.	Methoden	8
2.1.	Studiencharakteristik und Fragestellung	8
2.2.	Suchstrategie	10
2.3.	Ein- und Ausschluss der Studien.....	13
2.4.	Bewertung der Studien	14
2.5.	Datenextraktion	15
2.6.	Datenanalyse und Vergleich der Studien	21
3.	Resultate.....	21
3.1.	Methodologische Qualität der eingeschlossenen Studien	21
3.2.	Outcomes.....	23
3.2.1.	Relevante Reliabilität- und Validitätsparameter der einzelnen Schmerzprovokationstests ..	23
3.2.1.1.	Sensitivität.....	24
3.2.1.2.	Spezifität	24
3.2.1.3.	LR+	24
3.2.1.4.	LR-	24
3.2.1.5.	Positiv prädiktiver Wert	24
3.2.1.6.	Negativ prädiktiver Wert.....	24
3.2.1.7.	Interraterreliabilität	25

3.2.2.	Reliabilitäts- und Validitätsparameter der Clusters	25
3.2.2.1.	Arab et al. 2008.....	25
3.2.2.2.	Kokmeyer 2002.....	26
3.2.2.3.	Laslett 2005.....	27
3.2.2.4.	Laslett 2003.....	28
3.2.2.5.	Ozgocmen 2008	28
3.2.2.6.	Robinson et al. 2005.....	30
3.2.2.7.	van der Wurff et al. 2006	31
3.2.3.	Grafik zur Sensitivität aller daraufhin untersuchten Clusters	33
3.2.4.	Grafik zur Spezifität aller daraufhin untersuchten Clusters	34
3.2.5.	ROC plane mit allen Cluster-Werten	35
4.	Diskussion	39
4.1.	Stärken unserer Review.....	39
4.2.	Limitationen unserer Review	40
4.3.	Unterschiede zu bisher durchgeführten Reviews	41
4.4.	Interpretation der Resultate	42
4.5.	Relevanz für den Praktiker.....	44
4.6.	Bedeutung für die Forschung	45
5.	Schlussfolgerung	46
6.	Bibliographie	I
7.	Anhang	V
7.1.	Anhang I: Suchstrategie	V
7.2.	Anhang II: Definitionen der Outcomes	VII
7.3.	Anhang III: Testbeschreibungen	VIII
7.4.	Anhang IV: Quadas Tool	XII
7.5.	Anhang V: Tabellen der Einzeltests aus Studien	XIII

Abbildungsverzeichnis

Abbildung 1: Flussdiagramm unserer Suchstrategie	12
Abbildung 2: Zusammenfassung der methodologischen Qualität.....	22
Abbildung 3: Grafik zum Biasrisiko der eingeschlossenen Studien	23
Abbildung 4: Grafik der Sensitivität aller daraufhin untersuchten Clusters.....	33
Abbildung 5: Grafik der Spezifität aller daraufhin untersuchten Clusters	34
Abbildung 6: ROC plane mit Sensitivität im Verhältnis zu 1-Spezifität.....	35

Tabellenverzeichnis

Tabelle 1: Gründe für die Aussortierung.....	14
Tabelle 2: Datenextraktion	20
Tabelle 3: Beschreibung der Clusters	32
Tabelle 4: Resultatenextraktion	38
Tabelle 5: Definitionen der Outcomes.....	VII
Tabelle 6: Arab et al. 2008	XIII
Tabelle 7: Kokmeyer et al. 2002	XIV
Tabelle 8: Laslett et al. 2005	XV
Tabelle 9: Ozgocmen et al. 2008	XVI
Tabelle 10: Robinson et al. 2005	XVI

1. Einleitung

Das Clinical Reasoning spielt im physiotherapeutischen Prozess eine immer zentralere Rolle (A. Hüter-Becker & Dölken, 2005). Gemäss Art. 32 des schweizerischen Krankenversicherungsgesetzes müssen Behandlungsmassnahmen die Prinzipien der Wirksamkeit, Zweckmässigkeit und Wirtschaftlichkeit erfüllen (Oesch u. a., 2007). Clinical Reasoning kann etwa mit den deutschen Begriffen Begründen, Denken, Folgern, Argumentieren, Urteilen übersetzt werden. Sie drücken allerdings nur ein Teilverständnis des Clinical Reasonings aus, da es keine exakte Übersetzung für die deutsche Sprache gibt. Bis vor einigen Jahren stand beim Clinical Reasoning hauptsächlich die Interpretation, das Kognitive im Vordergrund. Es besteht aber die Tendenz zu einem vermehrt impliziten Clinical Reasoning, das heisst, es werden wieder vermehrt Gefühle und Intuition in den Prozess miteinbezogen (A. Hüter-Becker & Dölken, 2005).

Es gibt diverse Theorien über das Iliosakralgelenk (ISG) und seine Beweglichkeit. Zusätzlich kommen bei jedem Menschen unterschiedliche Kontextfaktoren hinzu, welche zu beachten sind. Daher gibt das Iliosakralgelenk häufig Anlass zu Diskussionen. Es ist schwierig, eine ISG-Problematik von einer Problematik mit ähnlichen Symptomen zu unterscheiden (Berthelot, Labat, Le Goff, Gouin, & Maugars, 2006a). In der Diagnostik und Therapie des ISG gibt es bis heute noch keinen Goldstandard (Beyerlein & Bessler, 2006), deshalb ist die Medizin auf sehr reliable und valide diagnostische Verfahren angewiesen (Pewsner u. a., 2001).

Um diese Untersuchungen durchführen zu können, werden von Ärzten und Forschern Hilfsmittel benötigt, die eine zusätzliche Kostenbelastung darstellen. Diese diagnostischen Verfahren bestehen aus Methoden wie Magnetresonanztomographie (MRI), Computertomographie (CT) oder Röntgen mit Injektionen von fluoreszierenden Kontrastmitteln (Mark Laslett, Charles N Aprill, Barry McDonald, & Young, 2005). Da solche Methoden, wie oben erwähnt, sehr kostspielig sind und zum Teil einer invasiven Anwendung bedürfen, stellen sie für den Patienten oft eine zusätzliche Belastung dar. Aufgrund dessen werden sie nur begrenzt durchgeführt.

In der Physiotherapie stehen solche aufwändige Apparaturen nicht zur Verfügung, deshalb werden Mobilitäts- und Schmerzprovokationstests durchgeführt, die je nach Ermessen des Physiotherapeuten evaluiert werden. Da es sehr schwierig ist, das ISG zuverlässig zu testen, gibt es aber häufig Abweichungen bei den Resultaten (M Laslett, B McDonald, Tropp, Ch Aprill, & Öberg, 2006).

Die Schwierigkeit besteht darin, die subjektiv wahrgenommenen Resultate zu objektivieren und anschliessend eine möglichst zuverlässige, evidenzbasierte und auf das Clinical Reasoning aufgebaute Diagnose zu stellen. Würde dies in Zukunft besser gelingen, könnten die Ärzte in ihrer Diagnosestellung durch die Physiotherapie unterstützt und bestätigt werden.

1.1. Theoretischer Hintergrund

1.1.1. Deskriptive und funktionelle Anatomie

Das Iliosakralgelenk (lat. articulatio iliosacralis) verbindet das Kreuzbein mit dem Darmbein. Die Gestalt des Sakrums und der Gelenkflächen ist abhängig von dem Wirbelsäulentyp. Bei stark ausgeprägten Krümmungen der Wirbelsäule ist das Sakrum fast horizontal ausgerichtet. Zusätzlich sind auch die Gelenkflächen deutlich gekrümmt und ausgeprägt konkav, was eine relativ grosse Beweglichkeit des ISG zur Folge hat. Bei den wenig ausgeprägten Krümmungen der Wirbelsäule ist dies anders: Hier ist das Sakrum beinahe vertikal. Die Gelenkflächen sind langgestreckt und kaum gekrümmt. Folglich ist das ISG nur gering beweglich (= typische Amphiarthrose). Die morphologischen Variationen des ISG sind beträchtlich (Kapandji, 2006).

Kurz gesagt ist das ISG relativ starr und wenig mobil, dennoch ist dieses Gelenk oft Ursache von Schmerzen in der Rücken-, Becken- und Beinregion (H. S. Robinson u. a., 2007).

Im Atlas „Funktionelle Anatomie“ von Ibrahim A. Kapandji werden zwei Bewegungen beschrieben, nämlich die Nutation und die Kontranutation. Unter dem Begriff Nutation versteht man die schraubige Bewegung nach vorne. Hierbei kippt das Promontorium nach vorne - unten und gleichzeitig gehen das Steissbein und die Kreuzbeinspitze nach hinten - oben. Bei der schraubigen Drehung nach hinten (Kontranutation) findet eine Verlagerung des Promontoriums nach oben - hinten statt, gleichzeitig bewegen sich das Steissbein und die Kreuzbeinspitze nach unten - vorne.

Verschiedene Autoren haben unterschiedliche Auffassungen bezüglich der Nutationsbewegung im ISG. Farabeuf beschreibt eine Kippbewegung im ISG, laut Bonnaire gibt es im ISG auch eine Kippbewegung, jedoch mit einem anderen Drehpunkt. Weisl hingegen beschreibt zwei Arten von Bewegungen: einerseits eine reine Translation und andererseits eine Kreisbewegung. Bei der reinen Translation gleitet das Kreuzbein längs der Achse des unteren Gelenkflächensegments. Die translatierte Distanz wird in Millimeter (mm) gemessen. Die Kreisbewegung ist eine Rotation um eine Achse, die sich ventrobasal vor dem Sakrum befindet, hier wird die Bewegung in Winkelgrad (°) gemessen (Kapandji, 2006). Die Rotation beträgt 1.2° bis 3° und die Translation 0.7 mm bis 2 mm (Berthelot et al., 2006).

Die anatomischen Unterschiede sind abhängig von Alter, Geschlecht, BMI und Geburtsvorgang. Es kann sein, dass es sogar Unterschiede der anatomischen Begebenheiten im Seitenvergleich gibt (Demir u. a., 2007). Die Beweglichkeit im ISG nimmt mit zunehmendem Alter ab (E. Cottant, 2004). Bewegungen mit sehr kleinem Ausmass sind extrem schwierig zu analysieren, dies merkt man auch aufgrund der diversen Konzepte dieser Theorien (Kapandji, 2006).

1.1.2. Ätiologie des ISG-Syndroms

Die Ursachen für ein ISG-Syndrom können vielfältig sein. Sie können grob in drei Bereiche eingeteilt werden: funktionelle Störungen, degenerative Veränderungen und entzündliche Veränderungen (Hildebrandt, Müller, & Pfingsten, 2005).

1.1.2.1. Funktionelle Störungen

Typische Ursachen für funktionelle Störungen sind etwa ein Verhebetrauma, ein „Tritt ins Leere“, beispielsweise beim Übersehen einer Treppenstufe, Hyperlaxität oder Blockierungen. Letztere können aber auch als Begleitblockierungen von Lendenwirbelsäulenerkrankungen oder anderen orthopädischen Erkrankungen auftreten (Hildebrandt u. a., 2005).

1.1.2.2. Degenerative Veränderungen

Degenerative Veränderungen treten hauptsächlich im Alter (Demir u. a., 2007) oder in Verbindung mit Brüchen oder Bandrupturen auf (Hildebrandt u. a., 2005).

1.1.2.3. Entzündliche Veränderungen

Entzündliche Veränderungen können einerseits isoliert, oder aufgrund von anderen entzündlichen Erkrankungen der Wirbelsäule, wie zum Beispiel Morbus Bechterew auftreten (Hildebrandt u. a., 2005).

1.1.3. Diagnostik des ISG

Ein Assessment ist ein Verfahren zur Beurteilung eines Patienten. Es werden systematisch aus therapeutischer Sicht relevante Eigenschaften evaluiert. Es gibt zwei Möglichkeiten, entweder findet eine Eigenevaluation oder eine Fremdevaluation statt, letztere ist objektiver. Das Assessmentergebnis kann teilweise in Zahlenwerten erfasst werden.

Um einen Überblick zu erhalten, wie funktionsfähig ein Patient ist oder wie die Behandlung wirkt, werden diverse Mess- und Testverfahren durchgeführt. Anhand der Resultate können individuell angepasste Ziele definiert werden. Zudem sind die Kontrolle des Verlaufes sowie die Messung der Behandlungsergebnisse möglich. Des Weiteren kann man standardisierte Assessments einsetzen, um eine Voraussage über das Ergebnis der Behandlung zu machen (Oesch u. a., 2007).

Wie schon erwähnt, gibt es Mobilitäts- und Schmerzprovokationstests, um das ISG zu diagnostizieren. Wir beschränken uns jedoch in unserer Review auf die Schmerzprovokationstests, da die Mobilitätstests eine schlechte Reliabilität und Validität haben. Um die Reliabilität und Validität dieser Tests herauszufinden, werden in den Studien invasive und/oder bildgebende Verfahren als Referenzstandard benutzt. Es gilt jedoch keines dieser diagnostischen Verfahren als Goldstandard (Beyerlein & Bessler, 2006).

1.1.3.1. Schmerzprovokationstests des ISG

Bei diesen Schmerzprovokationstests braucht man entweder die direkte Einwirkung auf die beiden Gelenkpartner oder man verwendet die indirekte Komponente über die unteren Extremitäten. Je nach Lagerung des Patienten kann sich der Therapeut die Schwerkraft zu Nutze machen oder er übt selber einen Druck aus (Trudelle, Gedda, & Gouilly, 2004).

Die Charakteristika der Tests erlauben es, diese in die Diagnostik zu integrieren. Durch die Kenntnis der Validität kann man sie bewusst wählen und kombinieren (E. Cottant, 2004). Es gibt Studien über die Reliabilität und Validität der einzelnen Schmerzprovokationstests und deren Kombinationen (Arab, Abdollahi, Joghataei, Golafshani, & Kazemnejad, 2009; Mark Laslett u. a., 2005; Ozgocmen, Bozgeyik, Kalcik, & Yildirim, 2008; H. S. Robinson u. a., 2007), andere Studien enthalten nur Testkombinationen einiger Schmerzprovokationstests (Mark Laslett, Young, Charles N Aprill, & Barry McDonald, 2003; Peter van der Wurff, Buijs, & Groen, 2006). Oesch et al. schreiben, dass die Clusters reliabler sind als die einzelnen Schmerzprovokationstests (Oesch u. a., 2007).

Ein Schmerzprovokationstest des ISG ist dann positiv, wenn die vertrauten Symptome während der Testdurchführung reproduziert oder verstärkt werden. Ein Test ist negativ, wenn die gleichen Symptome weder reproduziert noch verstärkt werden (Mark Laslett u. a., 2003).

Die Diagnostik dient zur Beurteilung, ob der Schmerz von den Gelenkstrukturen herrührt. Man benutzt den Schmerz als Verlaufszeichen sowie als Indikator für das Vorhandensein einer Problematik (Oesch u. a., 2007).

1.1.3.2. Invasive und bildgebende Verfahren des ISG

Neben den obengenannten Schmerzprovokationstests gibt es noch die invasiven und bildgebenden Verfahren. Diese bestehen aus Methoden wie Röntgen mit Injektionen von fluoreszierenden Kontrastmitteln (Mark Laslett u. a., 2005), Magnetresonanztomographie (MRI) und Computertomographie (CT) (Ozgocmen u. a., 2008).

1.1.4. Schmerz

„Schmerz ist ein unangenehmes Sinnes- und Gefühlserlebnis, das mit aktueller oder potentieller Gewebsschädigung verknüpft ist oder mit Begriffen einer solchen beschrieben wird.“
(Merskey, 1979)

Schmerz ist ein Warnsignal unseres Körpers. Durch mechanische, thermische oder chemische Reize wird das nozizeptive System aktiviert, das für die Detektion von gewebeschädigenden Reizen verantwortlich ist. Die Reizung erfolgt an den freien Nervenenden, die Information gelangt über das Rückenmark via afferente Bahnen (aufsteigende Bahnen) zum Zentralnervensystem (ZNS). Je nach Qualität und Dauer des Reizes wird Schmerz empfunden. Dies ist abhängig davon, ob die Reizschwelle überschritten wird oder nicht. Zudem ist Schmerz ein Symptom von Erkrankung oder zumindest ein Signal dafür, dass ein System im Körper dereguliert ist (van den Berg, 2003).

Die Auffassung und Definition bezüglich chronischer Schmerzen hat sich verändert. Für die Bezeichnung von chronischen beziehungsweise akuten Schmerzen ist nicht nur die Schmerzdauer, sondern hauptsächlich die Art des Schmerzes ausschlaggebend. Man spricht von chronischem oder chronifiziertem Schmerz, wenn nicht in erster Linie nozizeptive Vorgänge für den Schmerz verantwortlich sind. Oft stehen chronische Schmerzen in starkem Zusammenhang mit psychischen Aspekten. Folglich sind sie nicht nur mit Gewebeschädigungen verknüpft, sondern entwickeln ein „Eigenleben“. Chronische Erkrankungen, wie beispielsweise Arthrose, können aber sowohl eine akute als auch eine chronische Komponente enthalten (van den Berg, 2003).

Um eine klinische Diagnose von ISG-Schmerzen zu stellen, schlägt die International Association for the Study of Pain (IASP) ein Set von Schmerzkriterien vor, das sich auf die mechanischen Störungen dieses Gelenkes bezieht. Die Schmerzkriterien für ISG-Problematiken sehen wie folgt aus:

- 1) Vorhandener Schmerz im ISG-Bereich
- 2) Der vertraute Schmerz wird durch spezifische Schmerzprovokationstests reproduziert oder verstärkt
- 3) Die selektive Infiltration in das ISG erlöst den Patienten von seinen Schmerzen, wobei das lokale Anästhetikum in die Gelenkhöhle infiltriert wird (Mark Laslett, 2008)

Diese drei Kriterien werden mangels Informationen betreffend der Spezifität und Sensitivität kritisiert. Deswegen werden sie als nicht valide für die Diagnosestellung der ISG-Schmerzen angesehen. Wie man der Literatur entnehmen kann, scheint es so, dass bei schmerzenden ISG-Pathologien nicht nur ein intraartikuläres, sondern auch ein periartikuläres (Bänder und Muskeln) Problem besteht (Oesch u. a., 2007).

Mit der visuellen Analogeskala (VAS) wird der Schmerz beurteilt, der Dolometer reicht von 0 bis 10, wobei 0 keine Schmerzen und 10 unerträgliche Schmerzen bedeuten. Die VAS gilt als reliabel und valide (Oesch u. a., 2007).

1.1.5. Outcomes

Outcome bedeutet übersetzt Ergebnis, Resultat, Folge, Endergebnis sowie Endpunkt eines Behandlungsverlaufs.

Die Qualität der medizinischen Versorgung gewinnt stetig an Bedeutung, damit man den Patienten adäquat behandeln und gleichzeitig die Kosten senken kann. Solide Outcome-Informationen sind äusserst wichtige Bausteine für die evidenzbasierte Medizin (EBM) (Kane u. a., 2006).

Die Definitionen der für unsere Review relevanten Outcomes sind im Anhang II in einer Tabelle dargestellt.

1.2. Ziel

Die Reliabilität und Validität von Schmerzprovokationstests des ISG stellt bereits seit langer Zeit ein wichtiges Forschungsthema dar. Es ist bereits bekannt, dass Testkombinationen eine höhere Reliabilität und Validität haben als Einzeltests. Das Ziel unserer Review ist es daher, einen zusammenfassenden Überblick über die Reliabilität und Validität der verschiedenen Clusters zu geben.

2. Methoden

2.1. Studiencharakteristik und Fragestellung

Da es sich bei unserer Review um eine diagnostische Review handelt, waren die eingeschlossenen Studien Observationsstudien und diagnostische Studien mit folgenden Charakteristiken bezüglich Population, Intervention, Referenztest und Outcomes.

Population: Männer und Frauen ab 18 Jahren mit oder ohne ISG-Beschwerden

Die Intervention besteht aus folgenden Schmerzprovokationstests des ISG:

- Bilaterale Innenrotation der Hüfte
- Distractionstest
- Drop Test
- Gaenslen Test
- Kompressionstest
- Mennell-Test
- Patrick-Faber Test
- Resisted abduction Test
- Sacral thrust Test
- Thigh thrust Test
- Unilaterale Innenrotation der Hüfte

Da es unterschiedliche Nomenklaturen für die verschiedenen Tests gibt, haben wir, soweit möglich, im Anhang III die für uns am geläufigsten Synonyme aufgeführt.

Die Schmerzprovokationstests ermöglichen eine dichotome Testauswertung. Der Test ist entweder positiv oder negativ. Ein Testresultat ist positiv, wenn es eine Reproduktion oder Verstärkung des vertrauten Schmerzes gibt. Wenn keine Reproduktion oder Verstärkung des vertrauten Schmerzes erfolgt, gilt das Testresultat als negativ.

Die Schmerzprovokationstests wurden meist von Physiotherapeuten durchgeführt (unklar bei (Peter van der Wurff u. a., 2006); Studenten im letzten Ausbildungsjahr bei ((Kokmeyer, Van der Wurff, Aufdemkampe, & Fickenscher, 2002). Die bereits diplomierten Physiotherapeuten verfügen einerseits über eine Berufserfahrung von 1 Jahr (Arab u. a., 2009) bis zu 5.8 Jahre (H. S. Robinson u. a., 2007), oder andererseits über eine Spezialisierung in Wirbelsäulendiagnostik (Mark Laslett u. a., 2005; Mark Laslett u. a., 2003) oder in Manueller Therapie (H. S. Robinson u. a., 2007). Bei Ozgocmen et al. 08 führten die Tester eine vorhergehende Trainingssession zur Standardisierung der Testverfahren durch. In allen Studien wurden die Therapeuten weder über vorhergehende Testresultate noch über Patientendaten informiert. Die Therapeuten waren also diesbezüglich geblindet.

Die Referenztests bestanden entweder in einer Kontrastmittelinjektion (Mark Laslett u. a., 2005; Mark Laslett u. a., 2003; Peter van der Wurff u. a., 2006), eines MRI (Ozgocmen u. a., 2008) oder im Vergleich Tester 1 versus Tester 2 (Arab u. a., 2009; H. S. Robinson u. a., 2007).

Die Outcomes von Schmerzprovokationstests des ISG sind aus folgenden Werten aufgebaut:

- Sensitivität
- Spezifität
- Positiver prädiktiver Wert
- Negativer prädiktiver Wert
- Positive Likelihood Ratio
- Negative Likelihood Ratio

→ Intertesterreliabilität:

→ Kappa-Koeffizient

→ PABAK (prevalance-adjusted and bias-adjusted kappa)

Für die Messung der Schmerzintensität wurden die VAS verwendet.

2.2. Suchstrategie

Da keine Suche in nur einer Datenbank alle Studien identifiziert, haben wir in den folgenden Datenbanken gesucht: Cochrane, PubMed, Cinahl, Psycinfo und Web of Science. Wir haben uns in unserer Suche auf humanmedizinische Studien beschränkt. Dort wo es möglich war, haben wir Limits → Humans eingegeben, da oft auch Studien mit Tieren z.B. mit Pferden durchgeführt werden (Goff, Jeffcott, Jasiewicz, & McGowan, 2008). Falls diese Limitierung nicht möglich war, haben wir Begriffe wie ((humans) OR (population) OR (adults)) in unsere Suche integriert.

Wir haben unsere Suche in den fünf Datenbanken Cochrane, Pubmed, Cinahl, Psycinfo und Web of Science wie folgt aufgebaut:

Die 1. Suche führten wir mit den im Anhang I aufgelisteten Suchbegriffen durch, welche je nach Datenbank geringfügig variierten. Unsere 1. Suche ergab 381 Artikel, nach Überprüfung der Titel blieben noch 82 Artikel übrig.

Anschliessend haben wir noch eine 2. Suche basierend auf der ersten durchgeführt, indem wir die weiterführende Literatur zu jedem der 82 Artikel geprüft haben. Dies waren 1444 Artikel bzw. Texte.

Nach Überprüfen des Titels, des Abstracts sowie des Volltextes blieben von jenen 1444 Artikeln noch fünf übrig, welche wir in unsere Review einschlossen.

Unsere letzte elektronische Suche führten wir am 7. Mai 2010 durch. Dabei konnten wir noch zwei neue Studien einschliessen, welche unsere Ein- und Ausschlusskriterien erfüllten.

Bei der Suche sind wir einige Male auf den Namen Laslett gestossen. Einige seiner Artikel waren online nicht verfügbar, deshalb haben wir Dr. Mark Laslett kontaktiert um noch weitere unser Thema betreffende Informationen zu erhalten. Er hat uns eine Reihe von Artikeln

zugesandt, welche wir teils direkt für unsere Review und teils für den theoretischen Hintergrund nutzen konnten.

Bei der Handsuche sind wir auf die 31. Ausgabe der Zeitschrift „Kinésithérapie les annales“ gestossen und konnten den Inhalt für den theoretischen Hintergrund sowie für die Testabbildungen im Anhang III benutzen (E Cottant, 2004).

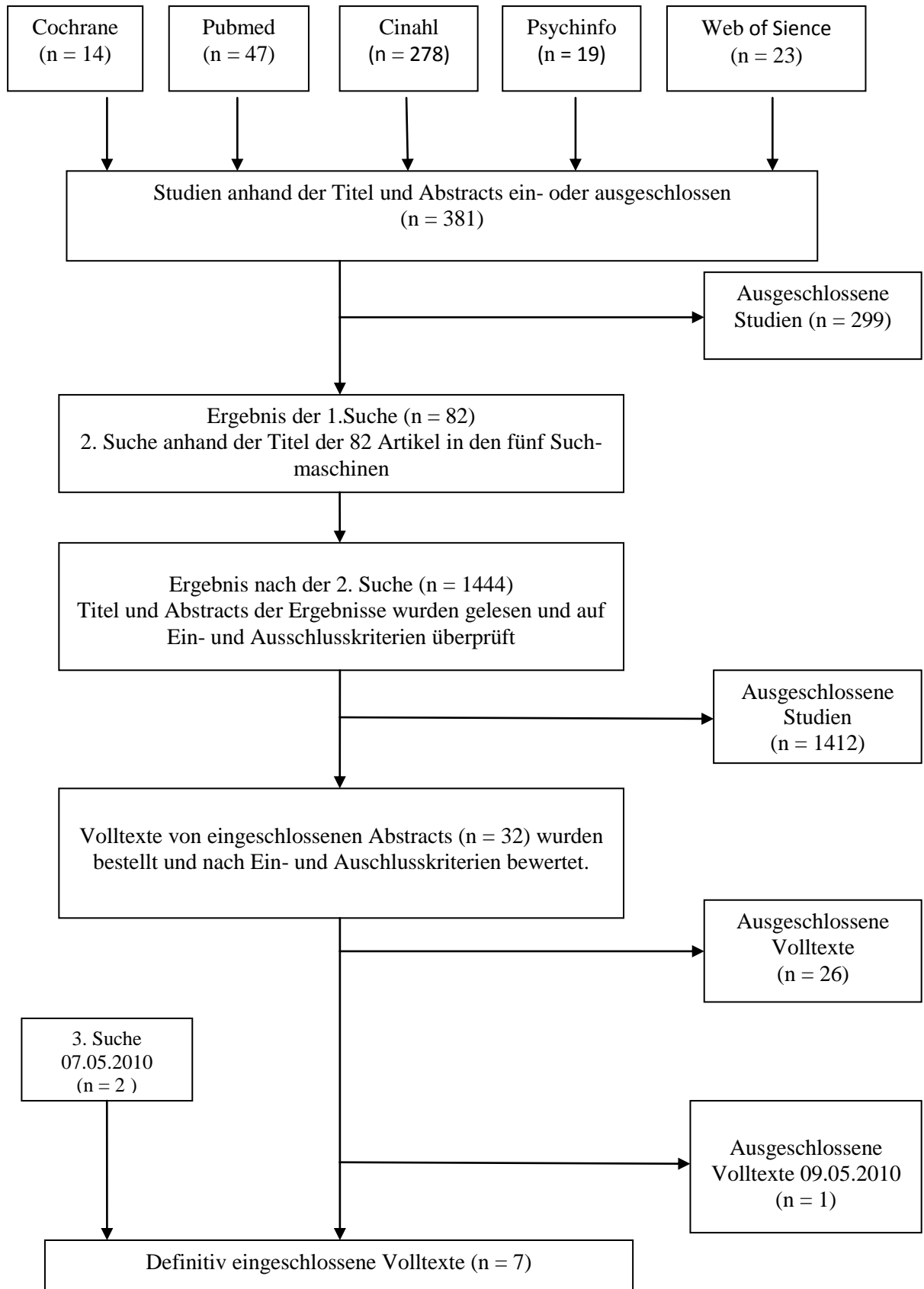


Abbildung 1: Flussdiagramm unserer Suchstrategie

2.3. Ein- und Ausschluss der Studien

Bei unserem Auswahlverfahren haben wir Studien anhand der unten genannten Ein- und Ausschlusskriterien aussortiert oder beibehalten.

Einschlusskriterien:

- Studien, welche einen oder mehrere ISG-Schmerzprovokationstests auf ihre Reliabilität und Validität überprüfen
- Studienpopulation bestehend aus erwachsenen Männern und Frauen (ab 18 Jahren)
- deutsche, französische oder englische Studien
- Observationstudien → musste evtl. im Volltext verifiziert werden, da diese Information nicht immer im Titel oder Abstrakt enthalten und ersichtlich ist. Dies bildete also kein absolutes Ausschlusskriterium für das erste Auswahlverfahren.

Ausschlusskriterien:

- Tiere, Veterinärmedizin
- Studien mit Kindern
- Case control studies
- Systematische Reviews und Metaanalysen

Unser Auswahlverfahren bestand aus zwei Hauptschritten:

1. Auswahl anhand von Titel und Abstrakt

Falls oben erwähnte Kriterien im Titel und Abstrakt erfüllt waren, wurden die Studien für den weiteren Verlauf eingeschlossen. Ebenso, falls im Titel und Abstrakt nicht klar ersichtlich war, ob die Studien die von uns gestellten Kriterien erfüllten. Es blieben 32 Artikel übrig.

2. Auswahl anhand der Volltexte

Ähnlich wie beim ersten Verfahren wurden nun die Volltexte auf die Ein- und Ausschlusskriterien überprüft.

Sieben Volltexte erfüllten unsere Kriterien und wurden eingeschlossen.

27 Volltexte erfüllten unsere Kriterien nicht und wurden aus unten erwähnten Gründen aussortiert:

Keine Observationsstudien	(20)
Case control study	(1)
Studie mit Kindern	(1)
Nicht gewählte Outcomes	(4)
Intervention bestand nicht aus Schmerzprovokationstests	(1)

Tabelle 1: Gründe für die Aussortierung

2.4. Bewertung der Studien

Da es sich bei den von uns eingeschlossenen Studien um diagnostische Studien handelt, verwendeten wir das Quadas-Tool (siehe Anhang IV) zur Bewertung der methodologischen Qualität der Studien. Das Quadas-Tool wurde entwickelt, damit eine evidenzbasierte Qualitätsbeurteilung bei Studien, die in eine diagnostische Review integriert werden möchten, möglich ist (Whiting, Rutjes, Reitsma, Bossuyt, & Kleijnen, 2003). Zudem wurden in einer weiteren Studie Brauchbarkeit und Validität untersucht, Whiting et al. beschreiben eine sehr gute Validität des Quadas-Tools (Whiting u. a., 2006). Die Schwierigkeit besteht allerdings darin, dass die zwei Reviewer, die unabhängig voneinander die Studie bewerten, nicht dasselbe unter den einzelnen Items verstehen könnten. Dies kann man jedoch umgehen, indem vorgängig zur Beurteilung der Qualität die einzelnen Punkte diskutiert werden um einen gemeinsamen Nenner zu finden.

Das Quadas-Tool enthält 14 Items, welche mit „Yes“, „Unclear“ oder „No“ bewertet werden können. Ein „Yes“ bedeutet ein geringes Biasrisiko, ein „Unclear“ steht für ein unklares und „No“ für ein hohes Biasrisiko. Whiting et al. beschreiben detailliert die einzelnen Items in ihrer Studie (Whiting u. a., 2003).

Die anhand des Volltextes eingeschlossenen Studien wurden separat von Christine und Leonie durchgelesen und bewertet. Anschliessend wurden die Resultate verglichen. Bei divergierenden Beurteilungen wurde diskutiert um eine jeweilige Entscheidung zu treffen. Falls dies nicht möglich war, wurde eine Drittperson hinzugezogen und um Rat gefragt.

2.5. Datenextraktion

Die extrahierten Daten der Studien (Studiencharakteristiken) sind in der nachfolgenden Tabelle ersichtlich. Die Extraktion soll helfen, einen besseren Überblick zu erhalten, um die Studien besser miteinander vergleichen zu können.

	Titel	Stichprobe	Ziel der Studie	Ein-/Ausschlusskriterien	Schmerzprovokationstests	Referenztest
Arab et al. 2008	Inter- and intra-examiner reliability of single and composites of selected motion palpation and pain provocation tests for sacroiliac joint	25 Teilnehmer (15 Männer, 10 Frauen) zwischen 20 und 65 Jahren	Die Studie testet gewählte Schmerzprovokationstests und Bewegungspalpatationstests und evaluiert die Inter- und Intratesterreliabilität von einzelnen und zusammengesetzten Tests.	<p>Einschlusskriterien:</p> <ul style="list-style-type: none"> - Schmerzen unterhalb L5 - Schmerzen rund ums ISG - Obere Region des ISG, um SIPS herum - Schmerzen an einer Gesässhälfte mit oder ohne Beinschmerzen <p>Ausschlusskriterien:</p> <ul style="list-style-type: none"> - Schmerzen nur in der Mitte oder symmetrisch oberhalb L5 - Ausstrahlender Schmerz mit neurologischen Zeichen - WS-OPs - Wirbelfrakturen - Beckenfraktur - UE-Fraktur - Spitalaufenthalt für schwerwiegende Traumen oder Autounfall - Beinlängendifferenz - Hüft-/Kniedysfunktion - Schwangerschaft - Innere Leiden (systemic disease) - Leber- und Nierenversagen 	<ul style="list-style-type: none"> - Patrick-Faber Test - Thigh thrust Test - Resisted abduction Test 	<ul style="list-style-type: none"> - Links-Rechts - Tester 1 - 2

	Titel	Stichprobe	Ziel der Studie	Ein-/Ausschlusskriterien	Schmerzprovokationstests	Referenztest
Kokmeyer et al. 2002	The reliability of multitest regimens with sacroiliac pain provocation tests	78 Teilnehmer (61 Männer, 17 Frauen zwischen 21 und 63 Jahren)	Berechnen der Interraterreliabilität durch das Benutzen von fünf Schmerzprovokationstests.	-	<ul style="list-style-type: none"> - Distractionstest - Kompressionstest - Tight thrust Test - Gaenslen Test - Patrick-Faber Test 	Tester 1-2
Laslett 2003	Diagnosing painful sacroiliac joints: A validity study of a McKenzie evaluation and sacroiliac provocation tests	48 Patienten (32 Frauen, 16 Männer)	Berechnung der diagnostischen Genauigkeit von Schmerzprovokationstests des Iliosakralgelenks	<p>Einschlusskriterien:</p> <ul style="list-style-type: none"> - Patienten mit Gesässschmerzen, mit oder ohne Symptomatik lumbal und in den Beinen - Gescheiterte therapeutische Behandlungen - Durchleben von bildgebenden Studien <p>Ausschlusskriterien:</p> <ul style="list-style-type: none"> - Zentrale oder symmetrische Schmerzen oberhalb von L5 - Klare Symptome einer Nervenkompression - Überweisung für andere Prozeduren als -Injektionen des ISG - Zu schwach für eine Ganzkörperuntersuchung - Ablehnen der Partizipation 	<ul style="list-style-type: none"> - Distractionstest - Thigh thrust Test - Gaenslen Test - Kompressionstest - Sacral thrust Testt 	Kontrastmittelinjektion

	Titel	Stichprobe	Ziel der Studie	Ein-/Ausschlusskriterien	Schmerzprovokationstests	Referenztest
Laslett 2005	Diagnosis of Sacroiliac Joint Pain: Validity of individual provocation tests and composites of tests	48 Patienten (32 Frauen, 16 Männer)	Diagnostische Aussagekraft von Schmerzprovokationstests des Iliosakralgelenks (einzeln oder in Kombination) in Bezug auf ein Standardkriterium	Einschlusskriterien: - Patienten mit Gesässschmerzen, mit oder ohne Symptomatik Lumbal und in den Beinen - Gescheiterte therapeutische Behandlungen - Durchleben von bildgebenden Studien Ausschlusskriterien: - Zentrale oder symmetrische Schmerzen oberhalb von L5 - Klare Symptome einer Nervenkompression - Überweisung für andere Prozeduren als -Injektionen des ISG - Zu schwach für eine Ganzkörperuntersuchung - Ablehnen der Partizipation	- Distractionstest - Thight thrust Test - Gaenslen Test - Kompressionstest - Sacral thrust Test	Kontrastmittelinjektion Initialinjektion: - Lidocaine Konfirmationsinjektion: - Bupivacaine

	Titel	Stichprobe	Ziel der Studie	Ein-/Ausschlusskriterien	Schmerzprovokationstests	Referenztest
Ozgocmen 2008	The value of sacroiliac pain provocation tests in early active sacroiliitis	40 Patienten zwischen 18 und 45 Jahren	Berechnung des Nutzen von Schmerzprovokationstests bei der Diagnose von Sacroiliitis	Einschlusskriterien: - Schmerzen Lumbal ≥ 40 mm (VAS 0-100) Ausschlusskriterien: - Neurologische Symptome in den Beinen - Straight leg-rising positiv - Mehr als 1,5 cm Beinlängendifferenz - Wadell-Score höher als 2 - Metabolische oder endokrine Beschwerden - akute Infektionen - Fibromyalgie - Schwangerschaft - Eingeschränkte oder schmerzhafte Hüftmobilität - Extreme Schmerzzustände - Nicht tolerieren körperlicher Untersuchung	- Kompressionstest - Distractionstest - Mennell-Test - Patrick-Faber Test - Thigh thrust Test - Sacral thrust Test	MRI
Robinson et al. 2005	The reliability of selected motion- and pain provocation tests for the sacroiliac joint	61 Teilnehmer (5 Männer, 56 Frauen) zwischen 18 und 50 Jahren Wobei: - 15 Teilnehmer spondylitis ankylosant - 30 Teilnehmer post-partum-Beckenschmerzen - 16 Teilnehmer ohne Rückenschmerzen	Testen der Inter-Tester-Reliabilität von einem Palpations- und sechs Schmerzprovokationstests für Schmerzen mit Ursprung im Iliosakral-Bereich.	Einschlusskriterien: - Ankylosierende Spondylitis (AS) auch ohne ankylosis in der Beckenregion und ohne sichtbare Kyphose - Frauen mit post-partum-Schmerzen > 6 Wochen - Kontrollgruppe: gesunde Teilnehmer ohne Rücken- Becken- oder Hüftschmerzen in den letzten 3 Monaten	- Kompressionstest - Distractionstest - Thigh thrust Test - Patrick-Faber Test - Bilaterale Innenrotation der Hüfte - Unilateral Innenrotation der Hüfte - Drop Test	Tester 1 - Tester 2

	Titel	Stichprobe	Ziel der Studie	Ein-/Ausschlusskriterien	Schmerzprovokationstests	Referenztest
van der Wurff et al. 2006	A Multitest Regimen of Pain Provocation Tests as an Aid to Reduce Unnecessary Minimally Invasive Sacroiliac Joint Procedures	60 Teilnehmer (13 Männer, 47 Frauen) zwischen 18 und 80 Jahren	Vergleichen der diagnostischen Genauigkeit von Multitest-Schemata mit fünf ISG-Schmerzprovokationstests.	<p>Einschlusskriterien:</p> <ul style="list-style-type: none"> - Schmerzen unterhalb L5 mit oder ohne Beinschmerzen - Schmerzen rund ums ISG über den Beckenkamm (unilateral) mit oder ohne Beinschmerzen - VAS \geq 45mm - Schmerz muss chronisch sein (>50d) <p>Ausschlusskriterien:</p> <ul style="list-style-type: none"> - Alter < 18 Jahre und > 80 Jahre - Spondylitis ankylosant - Beinlängendifferenz > 2 cm - Osteoporose - Tumore - Infektion und kürzliche Frakturen in der LWS-Region - Hüftosteoarthritis mit klinischen Symptomen - Rädikulärer Schmerz mit neurologischen Zeichen - Schwangerschaft - Patienten, welche Antikoagulantia einnehmen - Leber- und Nierenversagen - Nur Bandscheibenabnormalität mit Nervenwurzelkompression mit neurologischen Zeichen 	<ul style="list-style-type: none"> - Distractionstest - Kompressionstest - Thigh thrust Test - Patrick-Faber Test - Gaenslen Test 	Injektionstechnik nach Protokoll Schwarzer et al. (1ml Kontrastmittel wird injiziert, fluoroskopisch überprüfen)

Tabelle 2: Datenextraktion

2.6. Datenanalyse und Vergleich der Studien

Nachdem wir die für unsere Review relevanten Studien aussortiert hatten, haben wir aus den von uns eingeschlossenen Studien jene Daten extrahiert, welche die Reliabilität und Validität der Schmerzprovokationstests näher beschreiben. Die Studienauswertung mit dem Quadas-Tool ergaben einige Biasrisiken für einige der eingeschlossenen Studien. Wir haben sowohl die Outcomes für die einzelnen Schmerzprovokationstests wie auch für die Clusters extrahiert. Erstere haben wir allerdings nur im Anhang V detailliert aufgeführt. Bei der Analyse der in den Studien enthaltenen Outcomes haben wir uns hauptsächlich auf die Analyse der Clusters beschränkt, allerdings haben wir jeweils den besten und schlechtesten Test bezüglich der verschiedenen Outcomes kurz aufgeführt. Da einerseits die Clusters oder andererseits die Outcomes der einzelnen Studien bis auf eine Ausnahme (Mark Laslett u. a., 2005; Mark Laslett u. a., 2003) nicht homogen sind, d.h. mit einer Ausnahme sind nicht die gleichen Tests oder Outcomes enthalten, können wir diese nicht miteinander vergleichen. Demzufolge ist ein Pooling und die Darstellung der Daten in einem Forrest Plot nicht sinnvoll. Aufgrund dieser Sachlage haben wir uns entschieden, die Resultate qualitativ zu beschreiben.

3. Resultate

3.1. Methodologische Qualität der eingeschlossenen Studien

Die von uns eingeschlossenen Studien haben bis auf zwei eine sehr gute methodologische Qualität. Bei diesen zwei Studien ist das Biasrisiko leicht erhöht (Kokmeyer u. a., 2002; Ozgocmen u. a., 2008). Bei der Studie von Ozgocmen et al. ist unklar, ob der Referenzstandard ohne Kenntnis der Ergebnisse des Indextests durchgeführt worden ist (Frage 11). Des Weiteren war die Periode zwischen Index- und Referenztest nicht kurz genug, um ausschliessen zu können, dass sich die Ausgangsbedingungen des Patienten nicht verändert haben (Frage 4). Bei der Studie von Kokmeyer et al. ist ebenfalls unklar, wie gross die Zeitspanne zwischen dem 1. Test und dem 2. Test ist (Frage 4). Zusätzlich fehlen bei jener Studie die Ein- und Ausschlusskriterien (Frage 2).

Folgende Studien weisen eine sehr hohe Qualität und damit ein sehr geringes Biasrisiko auf: Arab et al 09; Laslett et al. 05; Laslett et al. 03; Robinson et al. 07; van der Wurff et al. 06.

In Abbildung 1 ist ersichtlich, welche Punkte des Quadas-Tools jeweils mit „ja“, „unklar“ oder „nein“ beantwortet wurden.

	Representative spectrum?	Acceptable reference standard?	Acceptable delay between tests?	Partial verification avoided?	Differential verification avoided?	Incorporation avoided?	Reference standard results blinded?	Index test results blinded?	Relevant clinical information?	Uninterpretable results reported?	Withdrawals explained?	Clearly selection criteria?	Possible replication index test?	Possible replication reference standard?
Arab et al. 2009	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Kokmeyer et al. 2002	+	+	?	+	+	+	+	+	+	+	+	-	+	+
Laslett et al. 2003	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Laslett et al. 2005	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Ozgocmen et al. 2008	+	+	-	+	+	+	?	+	+	+	+	+	+	+
Robinson et al. 2007	+	+	+	+	+	+	+	+	+	+	+	+	+	+
van der Wurff et al. 2006	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Abbildung 2: Zusammenfassung der methodologischen Qualität

Die Abbildung 2 gibt einen gesamthaften Überblick über das Biasrisiko bezüglich den von uns verwendeten Studien.

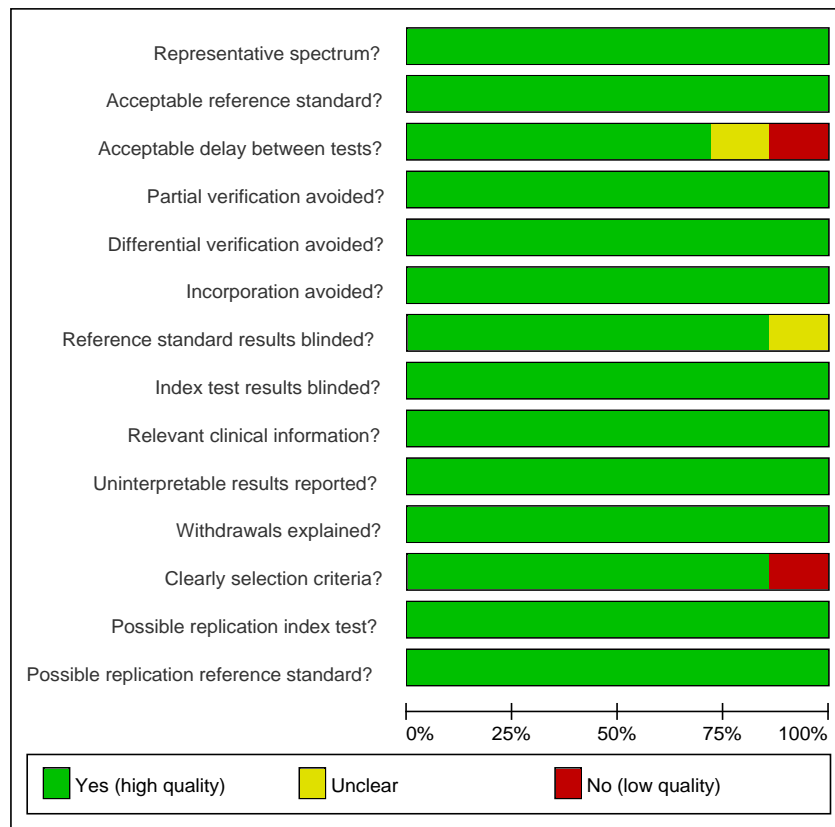


Abbildung 3: Grafik zum Biasrisiko der eingeschlossenen Studien

3.2. Outcomes

3.2.1. Relevante Reliabilität- und Validitätsparameter der einzelnen Schmerzprovokationstests

Da wir in unserer Review die Clusters anschauen, schneiden wir nur kurz die wichtigsten Randwerte der einzelnen Schmerzprovokationstests an. Aufgrund ähnlicher methodologischer Qualität haben wir für diese Resultatenextraktion jeweils die Studie mit der grössten Stichprobe und den nötigen Outcomes verwendet. Bei fortlaufendem Interesse bezüglich der Resultate der einzelnen Tests kann man im Anhang V ausführliche Tabellen der jeweiligen Studie konsultieren.

3.2.1.1. Sensitivität

Der Thigh thrust hat die beste Sensitivität mit einem Wert von 0.88.

Der Gaenslen Test (links) hat die schlechteste Sensitivität mit einem Wert von 0.50 (Mark Laslett u. a., 2005).

3.2.1.2. Spezifität

Der Distractionstest hat die beste Spezifität mit einem Wert von 0.81.

Der Kompressionstest und der Thigh thrust Test haben die schlechteste Spezifität mit einem Wert von 0.69 (Mark Laslett u. a., 2005).

3.2.1.3. LR+

Der Distractionstest hat die beste positive Likelihood Ratio mit einem Wert von 3.2.

Der Gaenslen Test (rechts) hat die niedrigste positive Likelihood Ratio mit einem Wert von 1.84 (Mark Laslett u. a., 2005).

3.2.1.4. LR-

Der Thigh thrust Test hat die beste negative Likelihood Ratio mit einem Wert von 0.18.

Der Gaenslen Test (rechts) hat die niedrigste negative Likelihood Ratio mit einem Wert von 0.66 (Mark Laslett u. a., 2005).

3.2.1.5. Positiv prädiktiver Wert

Der Distractionstest hat den besten positiven prädiktiven Wert mit 0.60.

Der Gaenslen Test (rechts) hat den schlechtesten positiven prädiktiven Wert mit 0.47 (Mark Laslett u. a., 2005).

3.2.1.6. Negativ prädiktiver Wert

Der Thigh thrust Test hat den besten negativen prädiktiven Wert mit 0.92.

Der Gaenslen Test (rechts) hat den schlechtesten negativen prädiktiven Wert mit 0.76 (Mark Laslett u. a., 2005).

3.2.1.7. Interraterreliabilität

Die Interraterreliabilität wird in den sieben eingeschlossenen Studien anhand des Kappa-Koeffizienten und des PABAK-Koeffizienten gemessen.

3.2.1.7.1. Kappa-Koeffizient

Der Thigh thrust Test hat den besten Kappa-Koeffizient mit 0.69.

Der Disktraktionstest hat den schlechtesten Kappa-Koeffizient mit 0.45 (Kokmeyer u. a., 2002).

3.2.1.7.2. PABAK- Koeffizient

Der Kompressionstest hat den besten PABAK- Koeffizient mit 0.87.

Der Disktraktionstest hat den schlechtesten PABAK- Koeffizient mit 0.44 (Kokmeyer u. a., 2002).

3.2.2. Reliabilitäts- und Validitätsparameter der Clusters

Wir haben die Beschreibung der Resultate der Clusters nach Studien aufgeteilt, um einen guten Überblick der extrahierten Resultate zu erhalten.

3.2.2.1. Arab et al. 2008

Arab et al. haben einen Cluster von drei Schmerzprovokationstests getestet. Die in den Clusters eingeschlossenen Tests sind Patrick-Faber Test, Thigh thrust Test und resisted abduction Test (Mac-Nab Test).

Für die Interraterreliabilität beträgt der Kappa-Koeffizient 0.59 mit einem 95%-Konfidenzintervall von 0.28-0.91 (rechts) und 0.51 mit einem 95%-Konfidenzintervall von 0.17-0.85 (links), wenn einer von drei Tests positiv ist. Wenn zwei von drei Tests positiv sind, beträgt der Kappa-Koeffizient 0.68 einem 95%-Konfidenzintervall von 0.35-1.01 (rechts) und 0.5 mit einem 95%-Konfidenzintervall von 0.06-0.95 (links). Wenn alle drei Tests positiv sind, misst der Kappa-Koeffizient 0.88 mit einem 95%-Konfidenzintervall von 0.66-1.10 (rechts) und 0.83 mit einem 95%-Konfidenzintervall von 0.51-1.15. Der Standardfehler vom Kappa-Koeffizient befindet sich zwischen 0.11 und 0.22.

Zusätzlich zum Kappa-Koeffizient wurde noch der PABAK-Koeffizient erforscht. Wenn einer von drei Tests positiv ist, beträgt der PABAK-Koeffizient 0.6 (rechts) und 0.52 (links). Wenn zwei von drei Tests positiv sind, ist der PABAK-Koeffizient 0.76 (rechts) und 0.68 (links). Der PABAK-Koeffizient beträgt 0.92 (beidseits), wenn alle drei Tests positiv sind (Arab u. a., 2009).

3.2.2.2. Kokmeyer 2002

Bei Kokmeyer et al. beinhaltet der Cluster folgende Tests: Distractionstest, Kompressions-test, Thigh thrust Test, Gaenslen Test, Patrick-Faber Test.

Wenn einer von fünf Tests positiv ist, beträgt der Kappa-Koeffizient 0.63 mit einem 95%-Konfidenzintervall von 0.47-0.83 für die ganze Stichprobe und 0.64 mit einem 95%-Konfidenzintervall von 0.44-0.84 für selektive Stichprobe (¹asymptomatische Subjekt wurden ausgeschlossen). Wenn zwei der fünf Tests positiv sind, beträgt der Kappa-Koeffizient 0.74 mit einem 95%-Konfidenzintervall von 0.54-0.94 für die ganze Stichprobe und 0.72 mit einem 95%-Konfidenzintervall von 0.51-0.93 für selektive Stichprobe (¹). Wenn drei von fünf Tests positiv sind, beträgt der Kappa-Koeffizient 0.70 mit einem 95%-Konfidenzintervall von 0.45-0.95 für die ganze Stichprobe und 0.69 mit einem 95%-Konfidenzintervall von 0.42-0.95 für selektive Stichprobe (¹). Wenn vier der fünf Tests positiv sind, beträgt der Kappa-Koeffizient 0.71 mit einem 95%-Konfidenzintervall von 0.38-1.03 für die ganze Stichprobe und 0.70 mit einem 95%-Konfidenzintervall von 0.37-1.03 für selektive Stichprobe (¹). Wenn alle fünf Tests positiv sind, beträgt der Kappa-Koeffizient 0.66 mit einem 95%-Konfidenzintervall von 0.00-1.32 für die ganze Stichprobe und 0.66 mit einem 95%-Konfidenzintervall von 0.00-1.32 für selektive Stichprobe (¹).

Wie bereits bei Arab et al. wurde zusätzlich zum Kappa-Koeffizient noch der PABAK-Koeffizient erforscht. Wenn einer von fünf Tests positiv ist, beträgt der PABAK-Koeffizient 0.65 für die ganze Stichprobe und 0.64 für selektive Stichprobe (¹). Wenn zwei von fünf Tests positiv sind, beträgt der PABAK-Koeffizient 0.85 für die ganze Stichprobe und 0.80 für selektive Stichprobe (¹). Wenn drei der fünf Tests positiv sind, beträgt der PABAK-Koeffizient 0.87 für die ganze Stichprobe und 0.83 für selektive Stichprobe (¹). Wenn vier von fünf Tests positiv sind, beträgt der PABAK-Koeffizient 0.92 für die ganze

Stichprobe und 0.90 für selektive Stichprobe (¹). Wenn alle fünf Tests positiv sind, beträgt der PABAK-Koeffizient 0.97 sowohl für die ganze als auch für die selektive (¹) Stichprobe.

3.2.2.3. Laslett 2005

Distraktionstest, Thigh thrust Test, Gaenslen Test (rechts und links), Kompressionstest, Sacral thrust Test bilden gemeinsam den Cluster bei Laslett et al. Es gibt fünf verschiedene Kompositionen: ein oder mehr positive Tests, zwei oder mehr positive Tests, drei oder mehr positive Tests, vier oder mehr positive Tests, fünf oder mehr positive Tests.

Bei der Komposition von einem oder mehr positiven Tests (≥ 1 positive Tests) beträgt die Sensitivität 1.00 mit einem 95%-Konfidenzintervall von 0.81-1.00, die Spezifität hat einen Wert von 0.44 mit einem 95%-Konfidenzintervall von 0.28-0.61. Der positiv prädiktive Wert beträgt 0.47 mit einem 95%-Konfidenzintervall von 0.32-0.63 und der negativ prädiktive Wert beträgt 1.00 mit einem 95%-Konfidenzintervall von 0.79-1.00. Zudem beträgt die positive Likelihood Ratio 1.78 mit einem 95%-Konfidenzintervall von 1.41-2.54, die negative Likelihood Ratio hat einen Wert von 0.00 mit einem 95%-Konfidenzintervall von 0.00-0.46.

Bei der zweiten Komposition (≥ 2 positive Tests) hat die Sensitivität einen Wert von 0.93 mit einem 95%-Konfidenzintervall von 0.72-0.99. Mit einem 95%-Konfidenzintervall von 0.48-0.80 beträgt die Spezifität 0.66. Der positiv prädiktive Wert hat einen Wert von 0.58 mit einem 95%-Konfidenzintervall von 0.39-0.75 und der negativ prädiktive Wert beträgt 0.96 mit einem 95%-Konfidenzintervall von 0.78-0.99. Bei dieser Komposition beträgt die positive Likelihood Ratio 2.73 mit einem 95%-Konfidenzintervall von 1.72-4.64, die negative Likelihood Ratio hat einen Wert von 0.10 mit einem 95%-Konfidenzintervall von 0.02-0.45.

Die dritte Komposition (≥ 3 positive Tests) hat eine Sensitivität von 0.94 mit einem 95%-Konfidenzintervall von 0.72-0.99 und eine Spezifität von 0.78 mit einem 95%-Konfidenzintervall von 0.61-0.89. Mit einem 95%-Konfidenzintervall von 0.47-0.84 beträgt der positiv prädiktive Wert 0.68 und der negativ prädiktive Wert beträgt 0.96 mit einem 95%-Konfidenzintervall von 0.81-0.99. Die positive Likelihood Ratio hat einen Wert von 4.29 mit einem 95%-Konfidenzintervall von 2.34-8.58 und mit einem 95%-Konfidenzintervall von 0.14-0.37 beträgt die negative Likelihood Ratio 0.80.

Bei der Komposition von vier oder mehr positiven Tests (≥ 4 positive Tests) beträgt die Sensitivität 0.60 mit einem 95%-Konfidenzintervall von 0.36-0.80, die Spezifität hat einen Wert von 0.81 mit einem 95%-Konfidenzintervall von 0.65-0.91. Der positiv prädiktive Wert beträgt 0.60 mit einem 95%-Konfidenzintervall von 0.36-0.80 und der negativ prädiktive Wert beträgt 0.81 mit einem 95%-Konfidenzintervall von 0.65-0.91. Mit einem 95%-Konfidenzintervall von 1.42-7.31 beträgt die positive Likelihood Ratio 3.20, die negative Likelihood Ratio hat einen Wert von 0.49 mit einem 95%-Konfidenzintervall von 0.24-0.83.

Bei der fünften Komposition (≥ 5 positive Tests) beträgt die Sensitivität 0.27 mit einem 95%-Konfidenzintervall von 0.11-0.52. Mit einem 95%-Konfidenzintervall von 0.72-0.95 beträgt die Spezifität 0.88. Der positiv prädiktive Wert hat einen Wert von 0.50 mit einem 95%-Konfidenzintervall von 0.22-0.79 und der negativ prädiktive Wert beträgt 0.72 mit einem 95%-Konfidenzintervall von 0.56-0.84. Die positive Likelihood Ratio hat einen Wert von 2.13 mit einem 95%-Konfidenzintervall von 0.64-6.83, die negative Likelihood Ratio beträgt 0.84 mit einem 95%-Konfidenzintervall von 0.54-1.11 (Mark Laslett u. a., 2005).

3.2.2.4. Laslett 2003

Bei Laslett et al. gibt es einen Cluster mit fünf Tests: Distractionstest, Thigh thrust Test, Gaenslen Test, Kompressionstest, Sacral thrust Test. Folgende Outcomes wurden für drei und mehr positive Tests berechnet. Für die Sensitivität ergibt sich ein Wert von 0.91 mit einem 95%-Konfidenzintervall von 0.62-0.98. Die Spezifität beträgt 0.87 mit einem 95%-Konfidenzintervall von 0.68-0.96. Die positive Likelihood Ratio ist 6.97 mit einem 95%-Konfidenzintervall von 2.70-20.27. Die negative Likelihood Ratio beträgt 0.11 mit einem 95%-Konfidenzintervall von 0.02-0.44 (Mark Laslett u. a., 2003).

3.2.2.5. Ozgocmen 2008

Ozgocmen et al. haben zwei verschiedene Clusters, der eine Cluster besteht aus drei und der andere Cluster aus fünf Schmerzprovokationstests.

Gaenslen Test, Mennell Test und Thigh thrust Test bilden gemeinsam einen Cluster. Die folgenden Werte gelten, wenn zwei der drei Tests positiv sind. Die Sensitivität beträgt 0.55 mit einem 95%-Konfidenzintervall von 0.22-0.84 (rechts) und 0.45 mit einem 95%-Konfidenzintervall von 0.18-0.75 (links). Die Spezifität hat einen Wert von 0.83 mit einem 95%-Konfidenzintervall von 0.65-0.93 (rechts) und 0.86 mit einem 95%-Konfidenzintervall von 0.67-0.95 (links). Der positiv prädiktive Wert beträgt 0.50 mit einem 95%-Konfidenzintervall von 0.20-0.79 (rechts) und 0.55 mit einem 95%-Konfidenzintervall von 0.22-0.84 (links). Der negativ prädiktive Wert beträgt 0.86 mit einem 95%-Konfidenzintervall von 0.68-0.95 (rechts) und 0.80 mit einem 95%-Konfidenzintervall von 0.61-0.91. Die positive Likelihood Ratio hat einen Wert von 3.44 mit einem 95%-Konfidenzintervall von 1.27-9.29 (rechts) und 3.29 mit einem 95%-Konfidenzintervall von 1.07-10.0. Die negative Likelihood Ratio beträgt 0.52 mit einem 95%-Konfidenzintervall von 0.25-1.11 (rechts) und 0.63 mit einem 95%-Konfidenzintervall von 0.36-1.09. Der Kappa-Koeffizient beträgt 0.47 mit einem 95%-Konfidenzintervall von 0.15-0.78 (rechts) und 0.43 mit einem 95%-Konfidenzintervall von 0.09-0.76 (links).

Der zweite Cluster (eine Komposition bestehend aus fünf Tests) enthält den Gaenslen Test Patrick-Faber Test, Mennell-Test, Thigh thrust Test und der Sacral thrust Test. Die nachfolgenden Werte gelten, wenn drei der fünf Tests positiv sind. Die Sensitivität hat einen Wert von 0.43 mit einem 95%-Konfidenzintervall von 0.15-0.77 (rechts) und 0.45 mit einem 95%-Konfidenzintervall von 0.18-0.75 (links). Die Spezifität beträgt 0.83 mit einem 95%-Konfidenzintervall von 0.65-0.93 (rechts) und 0.89 mit einem 95%-Konfidenzintervall von 0.71-0.97 (links). Der positiv prädiktive Wert beträgt 0.44 mit einem 95%-Konfidenzintervall von 0.15-0.77 (rechts) und 0.62 mit einem 95%-Konfidenzintervall von 0.25-0.89 (links). Der negativ prädiktive Wert beträgt 0.83 mit einem 95%-Konfidenzintervall von 0.65-0.93 (rechts) und 0.81 mit einem 95%-Konfidenzintervall von 0.62-0.92. Die positive Likelihood Ratio hat einen Wert von 2.75 mit einem 95%-Konfidenzintervall von 0.93-8.15 (rechts) und 4.39 mit einem 95%-Konfidenzintervall von 1.25-15.36. Die negative Likelihood Ratio beträgt 0.66 mit einem 95%-Konfidenzintervall von 0.36-1.20 (rechts) und 0.60 mit einem 95%-Konfidenzintervall von 0.35-1.05. Der Kappa-Koeffizient beträgt 0.60 mit einem 95%-

Konfidenzintervall von 0.31-0.88 (rechts) und 0.53 mit einem 95%-Konfidenzintervall von 0.20-0.86 (links).

Die folgenden Werte gelten, wenn vier der fünf Tests positiv sind. Die Sensitivität hat einen Wert von 0.45 mit einem 95%-Konfidenzintervall von 0.15-0.78 (rechts) und 0.45 mit einem 95%-Konfidenzintervall von 0.18-0.75 (links). Die Spezifität beträgt 0.84 mit einem 95%-Konfidenzintervall von 0.66-0.94 (rechts) und 0.93 mit einem 95%-Konfidenzintervall von 0.75-0.98 (links). Der positiv prädiktive Wert beträgt 0.45 mit einem 95%-Konfidenzintervall von 0.15-0.78 (rechts) und 0.71 mit einem 95%-Konfidenzintervall von 0.30-0.94 (links). Der negativ prädiktive Wert beträgt 0.84 mit einem 95%-Konfidenzintervall von 0.66-0.94 (rechts) und 0.81 mit einem 95%-Konfidenzintervall von 0.63-0.92. Die positive Likelihood Ratio hat einen Wert von 2.75 mit einem 95%-Konfidenzintervall von 0.93-8.16 (rechts) und 6.6 mit einem 95%-Konfidenzintervall von 1.49-29.11. Die negative Likelihood Ratio beträgt 0.66 mit einem 95%-Konfidenzintervall von 0.36-1.20 (rechts) und 0.58 mit einem 95%-Konfidenzintervall von 0.33-1.0. Der Kappa-Koeffizient beträgt 0.51 mit einem 95%-Konfidenzintervall von 0.20-0.82 (rechts) und 0.54 mit einem 95%-Konfidenzintervall von 0.18-0.89 (links) (Ozgocmen u. a., 2008).

3.2.2.6. Robinson et al. 2005

Robinson et al. beschreibt zwei verschiedene Clusters.

Distraktionstest, Thigh thrust Test, Patrick-Faber Test, bilaterale Innenrotation der Hüfte und unilaterale Innenrotation der Hüfte bilden den ersten Cluster. Wenn drei von fünf Tests positiv sind, beträgt der Kappa-Koeffizient 0.60 mit einem 95%-Konfidenzintervall von 0.40-0.80 (rechts), 0.69 mit einem 95%-Konfidenzintervall von 0.51-0.87 (links) für die ganze Stichprobe und 0.51 mit einem 95%-Konfidenzintervall von 0.26-0.76 (rechts), 0.60 mit einem 95%-Konfidenzintervall von 0.36-0.84 (links) für die selektive Stichprobe (nur Patienten). Wenn vier von fünf Tests positiv sind, beträgt der Kappa-Koeffizient 0.71 mit einem 95%-Konfidenzintervall von 0.49-0.93 (rechts), 0.75 mit einem 95%-Konfidenzintervall von 0.56-0.94 (links) für die ganze Stichprobe und 0.68 mit einem 95%-Konfidenzintervall von 0.44-0.92 (rechts), 0.71 mit einem 95%-Konfidenzintervall von 0.49-0.93 (links) für die selektive Stichprobe (nur Patienten).

In dem zweiten Cluster sind folgende Tests enthalten: Thigh thrust Test, Patrick-Faber Test und Innenrotation der Hüfte. Wenn zwei der drei Tests positiv sind, beträgt der Kappa-Koeffizient 0.71 mit einem 95%-Konfidenzintervall von 0.47-0.95 (rechts), 0.75 mit einem 95%-Konfidenzintervall von 0.54-0.96 (links) für die ganze Stichprobe und 0.69 mit einem 95%-Konfidenzintervall von 0.44-0.94 (rechts), 0.72 mit einem 95%-Konfidenzintervall von 0.49-0.95 (links) für die selektive Stichprobe (nur Patienten) (H. S. Robinson u. a., 2007).

3.2.2.7. van der Wurff et al. 2006

Distraktionstest, Kompressionstest, Thigh thrust Test, Patrick-Faber Test, Gaenslen Test bilden gemeinsam den Cluster bei van der Wurff et al.

Die folgenden Resultate gelten, wenn drei und mehr Tests positiv sind. Die Sensitivität beträgt 0.85 mit einem 95%-Konfidenzintervall von 0.72-0.99, die Spezifität hat einen Wert von 0.79 mit einem 95%-Konfidenzintervall von 0.65-0.93. Mit einem 95%-Konfidenzintervall von 0.62-0.92 beträgt der positiv prädiktiver Wert 0.77, der negativ prädiktiver Wert beträgt 0.87 mit einem 95%-Konfidenzintervall von 0.74-0.99. Die positive Likelihood Ratio hat einen Wert von 4.02 mit einem 95%-Konfidenzintervall von 2.04-7.89 und die negative Likelihood Ratio beträgt 0.19 mit einem 95%-Konfidenzintervall von 0.07-0.47. Die maximale area under the curve (AUC) beträgt 0.799, wenn genau drei der fünf Tests positiv sind (Peter van der Wurff u. a., 2006).

	Arab et al. 2008	Kokmeyer et al. 2002	Laslett et al. 2003	Laslett et al. 2005	Ozgocmen et al 2008	Robinson et al. 2005	Van der Wurff 06
Cluster 1	<ul style="list-style-type: none"> - Patrick-Faber Test - Thigh thrust Test - Resisted abduction Test 	<ul style="list-style-type: none"> - Distractionstest - Kompressionstest - Thigh thrust Test - Patrick-Faber Test - Gaenslen Test 	<ul style="list-style-type: none"> - Distractionstest - Thigh thrust Test - Gaenslen Test - Kompressionstest - Sacral thrust Test 	<ul style="list-style-type: none"> - Distractionstest - Thigh thrust Test - Gaenslen Test (rechts und links) - Kompressionstest - Sacral thrust Test 	<ul style="list-style-type: none"> - Gaenslen Test - Mennell-Test - Thigh thrust Test 	<ul style="list-style-type: none"> - Distractionstest - Thigh thrust Test - Patrick-Faber Test - Bilaterale Innenrotation der Hüfte - Unilaterale Innenrotation der Hüfte 	<ul style="list-style-type: none"> - Distractionstest - Kompressionstest - Thigh thrust Test - Patrick-Faber Test - Gaenslen Test
Cluster 2	-	-	-	-	<ul style="list-style-type: none"> - Gaenslen Test - Patrick-Faber Test - Mennell-Test - Thigh thrust Test - Sacral thrust Test 	<ul style="list-style-type: none"> - Thigh thrust Test - Patrick-Faber Test - Innenrotation der Hüfte 	-

Tabelle 3: Beschreibung der Clusters

3.2.3. Grafik zur Sensitivität aller daraufhin untersuchten Clusters

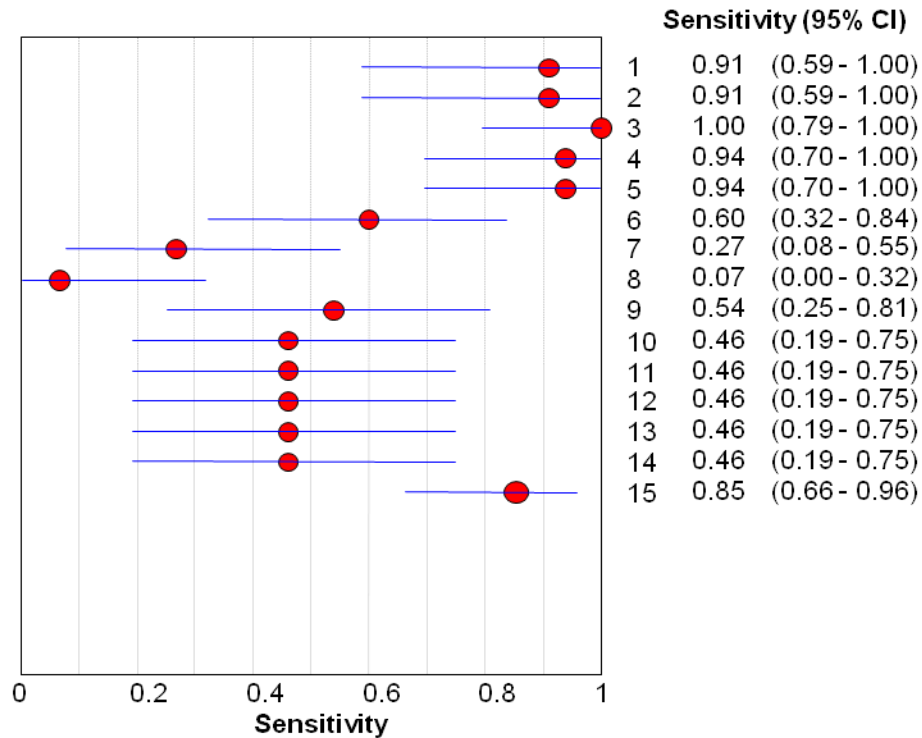


Abbildung 4: Grafik der Sensitivität aller daraufhin untersuchten Clusters

Legende:

- 1: Laslett et al. 2003 n=43
- 2: Laslett et al. 2003 n=34
- 3: Laslett et al. 2005 (≥ 1 positive Tests)
- 4: Laslett et al. 2005 (≥ 2 positive Tests)
- 5: Laslett et al. 2005 (≥ 3 positive Tests)
- 6: Laslett et al. 2005 (≥ 4 positive Tests)
- 7: Laslett et al. 2005 (≥ 5 positive Tests)
- 8: Laslett et al. 2005 (≥ 6 positive Tests)
- 9: Ozgocmen et al. 2008 (2 von 3 Tests positiv, rechts)
- 10: Ozgocmen et al. 2008 (2 von 3 Tests positiv, links)
- 11: Ozgocmen et al. 2008 (3 von 5 Tests positiv, rechts)
- 12: Ozgocmen et al. 2008 (3 von 5 Tests positiv, links)
- 13: Ozgocmen et al. 2008 (4 von 5 Tests positiv, rechts)
- 14: Ozgocmen et al. 2008 (4 von 5 Tests positiv, links)
- 15: van der Wurff et al. 2006 (≥ 3 positive Tests)

Die Abbildung 4 zeigt die Sensitivität der daraufhin untersuchten Clusters. Der rote Punkt widerspiegelt die Stichprobengrösse. Die Länge der Linie gibt Auskunft über die Grösse des 95%-Konfidenzintervalls.

3.2.4. Grafik zur Spezifität aller daraufhin untersuchten Clusters

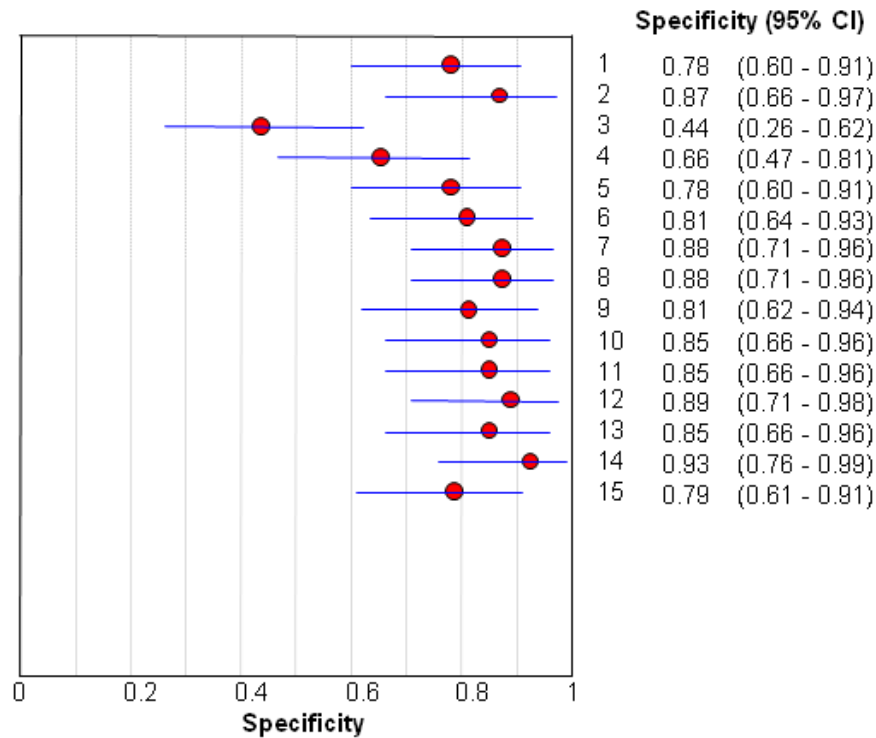


Abbildung 5: Grafik der Spezifität aller daraufhin untersuchten Clusters

Legende:

- 1: Laslett et al. 2003 n=43
- 2: Laslett et al. 2003 n=34
- 3: Laslett et al. 2005 (≥ 1 positive Tests)
- 4: Laslett et al. 2005 (≥ 2 positive Tests)
- 5: Laslett et al. 2005 (≥ 3 positive Tests)
- 6: Laslett et al. 2005 (≥ 4 positive Tests)
- 7: Laslett et al. 2005 (≥ 5 positive Tests)
- 8: Laslett et al. 2005 (≥ 6 positive Tests)
- 9: Ozgocmen et al. 2008 (2 von 3 Tests positiv, rechts)
- 10: Ozgocmen et al. 2008 (2 von 3 Tests positiv, links)
- 11: Ozgocmen et al. 2008 (3 von 5 Tests positiv, rechts)
- 12: Ozgocmen et al. 2008 (3 von 5 Tests positiv, links)
- 13: Ozgocmen et al. 2008 (4 von 5 Tests positiv, rechts)
- 14: Ozgocmen et al. 2008 (4 von 5 Tests positiv, links)
- 15: van der Wurff et al. 2006 (≥ 3 positive Tests)

Die Abbildung 5 zeigt die Spezifität der daraufhin untersuchten Clusters. Der rote Punkt widerspiegelt die Stichprobengrösse. Die Länge der Linie gibt Auskunft über die Grösse des 95%-Konfidenzintervalls.

3.2.5. ROC plane mit allen Cluster-Werten

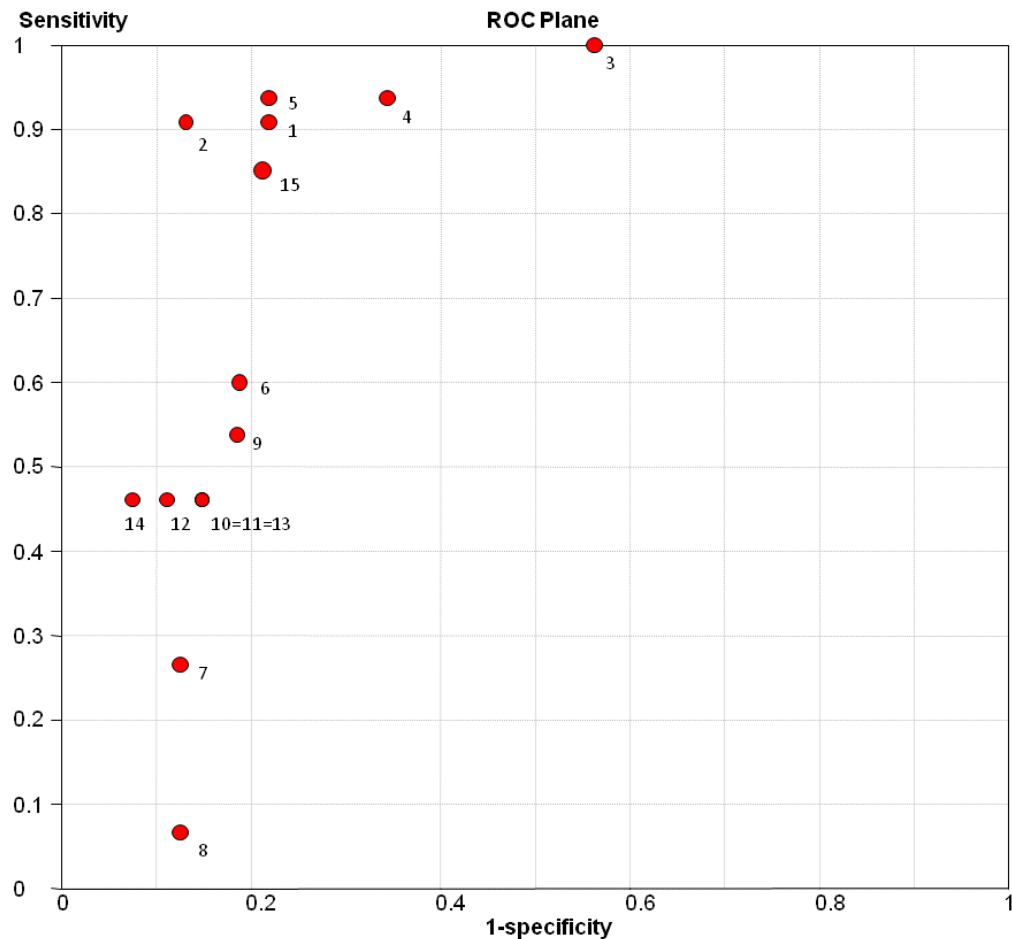


Abbildung 6: ROC plane mit Sensitivität im Verhältnis zu 1-Spezifität

Legende:

- 1: Laslett et al. 2003 n=43
- 2: Laslett et al. 2003 n=34
- 3: Laslett et al. 2005 (≥ 1 positive Tests)
- 4: Laslett et al. 2005 (≥ 2 positive Tests)
- 5: Laslett et al. 2005 (≥ 3 positive Tests)
- 6: Laslett et al. 2005 (≥ 4 positive Tests)
- 7: Laslett et al. 2005 (≥ 5 positive Tests)
- 8: Laslett et al. 2005 (≥ 6 positive Tests)
- 9: Ozgocmen et al. 2008 (2 von 3 Tests positiv, rechts)
- 10: Ozgocmen et al. 2008 (2 von 3 Tests positiv, links)
- 11: Ozgocmen et al. 2008 (3 von 5 Tests positiv, rechts)
- 12: Ozgocmen et al. 2008 (3 von 5 Tests positive, links)
- 13: Ozgocmen et al. 2008 (4 von 5 Tests positiv, rechts)
- 14: Ozgocmen et al. 2008 (4 von 5 Tests positiv, links)
- 15: van der Wurff et al. 2006 (≥ 3 positive Tests)

Die Abbildung 6 zeigt die Sensitivität im Verhältnis zu 1-Spezifität. Das Resultat ist umso besser, je mehr es sich der oberen linken Ecke nähert.

	Sensitivität mit 95% CI	Spezifität mit 95% CI	Positiver prädiktiver Wert mit 95% CI	Negativer prädiktiver Wert mit 95% CI	Positive Likelihood Ratio mit 95% CI	Negative Likelihood Ratio mit 95% CI	Intertesterreliabilität Kappa-Koeffizient (SE) mit 95% CI	PABAK (R, L)
Arab et al. 2008								
1 von 3 Tests pos.								
- rechts	-	-	-	-	-	-	0.59 (0.16) (0.28-0.91)	0.6
- links	-	-	-	-	-	-	0.51 (0.17) (0.17-0.85)	0.52
2 von 3 Tests pos.								
- rechts	-	-	-	-	-	-	0.68 (0.16) (0.35-1.01)	0.76
- links	-	-	-	-	-	-	0.5 (0.22) (0.06-0.95)	0.68
3 von 3 Tests pos.								
- rechts	-	-	-	-	-	-	0.88 (0.11) (0.60-1.10)	0.92
- links	-	-	-	-	-	-	0.83 (0.16) (0.51-1.15)	0.92
Kokmeyer et al. 2002								
1 von 5 Tests pos.	-	-	-	-	-	-	0.63 (0.093) (0.47-0.83)	0.65
1 von 5 Tests pos. *	-	-	-	-	-	-	0.64 (0.10) (0.44-0.84)	0.64
2 von 5 Tests pos.	-	-	-	-	-	-	0.74 (0.10) (0.54-0.94)	0.85
2 von 5 Tests pos. *	-	-	-	-	-	-	0.72 (0.11) (0.51-0.93)	0.80
3 von 5 Tests pos.	-	-	-	-	-	-	0.70 (0.13) (0.45-0.95)	0.87
3 von 5 Tests pos. *	-	-	-	-	-	-	0.69 (0.13) (0.42-0.95)	0.83
4 von 5 Tests pos.	-	-	-	-	-	-	0.71 (0.17) (0.38-1.03)	0.92
4 von 5 Tests pos. *	-	-	-	-	-	-	0.70 (0.17) (0.37-1.03)	0.90
5 von 5 Tests pos.	-	-	-	-	-	-	0.66 (0.34) (0.00-1.32)	0.97
5 von 5 Tests pos. *	-	-	-	-	-	-	0.66 (0.34) (0.00-1.32)	0.97
* exklusiv asymptomatische Patienten								

	Sensitivität mit 95% CI	Spezifität mit 95% CI	Positiver prädiktiver Wert mit 95% CI	Negativer prädiktiver Wert mit 95% CI	Positive Likelihood Ratio mit 95% CI	Negative Likelihood Ratio mit 95% CI	Intertesterreliabilität Kappa-Koeffizient (SE) mit 95% CI	PABAK (R, L)
Laslett 2003								
≥3 positive Tests pos. (n=48)	0.91 (0.62 - 0.98)	0.78 (0.61 - 0.89)	-	-	4.16 (2.16 - 8.39)	0.12 (0.02 - 0.49)	-	-
≥3 positive Tests pos. (n=34;-14 nach erstem Testverfahren, starke Veränderungen der Symptome)	0.91 (0.62 - 0.98)	0.87 (0.68 - 0.96)	-	-	6.97 (2.7 - 20.27)	0.11 (0.02 - 0.44)	-	-
Laslett 2005								
≥ 1 positiver Test	1.00 (0.81 - 1.00)	0.44 (0.28 - 0.61)	0.47 (0.32 - 0.63)	1.00 (0.79 - 1.00)	1.78 (1.41 - 2.54)	0.00 (0.00 - 0.46)	-	-
≥ 2 positive Tests	0.93 (0.72 - 0.99)	0.66 (0.48 - 0.80)	0.58 (0.39 - 0.75)	0.96 (0.78 - 0.99)	2.73 (1.72 - 4.64)	0.10 (0.02 - 0.45)	-	-
≥ 3 positive Tests	0.94 (0.72 - 0.99)	0.78 (0.61 - 0.89)	0.68 (0.47 - 0.84)	0.96 (0.81 - 0.99)	4.29 (2.34 - 8.58)	0.80 (0.14 - 0.37)	-	-
≥ 4 positive Tests	0.60 (0.36 - 0.80)	0.81 (0.65 - 0.91)	0.60 (0.36 - 0.80)	0.81 (0.65 - 0.91)	3.20 (1.42 - 7.31)	0.49 (0.24 - 0.83)	-	-
≥ 5 positive Tests	0.27 (0.11 - 0.52)	0.88 (0.72 - 0.95)	0.50 (0.22 - 0.79)	0.72 (0.56 - 0.84)	2.13 (0.64 - 6.83)	0.84 (0.54 - 1.11)	-	-
Ozgocmen 2008								
2 von 3 Tests pos.								
- rechts	0.55 (0.22 - 0.84)	0.83 (0.65 - 0.93)	0.50 (0.20 - 0.79)	0.86 (0.68 - 0.95)	3.44 (1.27 - 9.29)	-	0.47 (0.15 - 0.78)	-
- links	0.45 (0.18 - 0.75)	0.86 (0.67 - 0.95)	0.55 (0.22 - 0.84)	0.80 (0.61 - 0.91)	3.29 (1.07 - 10.0)	-	0.43 (0.09 - 0.76)	-
3 von 5 Tests pos.								
- rechts	0.43 (0.15 - 0.77)	0.83 (0.65 - 0.93)	0.44 (0.15 - 0.77)	0.83 (0.65 - 0.93)	2.75 (0.93 - 8.15)	-	0.60 (0.31 - 0.88)	-
- links	0.45 (0.18 - 0.75)	0.89 (0.71 - 0.97)	0.62 (0.25 - 0.89)	0.81 (0.62 - 0.92)	4.39 (1.25 - 15.36)	-	0.53 (0.20 - 0.86)	-
4 von 5 Tests pos.								
- rechts	0.45 (0.15 - 0.78)	0.84 (0.66 - 0.94)	0.45 (0.15 - 0.78)	0.84 (0.66 - 0.94)	2.75 (0.93 - 8.16)	-	0.51 (0.20 - 0.82)	-
- links	0.45 (0.18 - 0.75)	0.93 (0.75 - 0.98)	0.71 (0.30 - 0.94)	0.81 (0.63 - 0.92)	6.6 (1.49 - 29.11)	-	0.54 (0.18 - 0.89)	-

4. Diskussion

In den eingeschlossenen Studien wurden Schmerzprovokationstests auf ihre Validität und/oder Reliabilität überprüft. Clusters sind valider und reliabler als einzelne Schmerzprovokationstests. Jedoch kann aufgrund unserer extrahierten Resultate kein Cluster empfohlen werden, da keines bezüglich allen Outcomes überragend ist.

4.1. Stärken unserer Review

Ganz allgemein gesehen können wir bezüglich unserer Review einige Punkte erwähnen, welche wir als Stärken unserer Review ansehen. Wir verwendeten für die Bewertung der Studienqualität das Quadas-Tool, welches sich wie im Methodenteil bereits erwähnt, sehr gut eignet, da die Validität für die Bewertung diagnostischer Studien sehr gut ist. (Whiting u. a., 2006). Die Bewertung unserer Studien mit dem Quadas-Tool ergab wie in Abbildung 1 ersichtlich ist eine sehr gute Methodologische Qualität und folglich ein geringes Biasrisiko für die eingeschlossenen Studien. Aufgrund dessen ist die Wahrscheinlichkeit von fehlerhaften Resultaten geringer und die Resultate weisen dadurch eine höhere Glaubwürdigkeit auf. Trotz der guten Qualität, welche sich durch die Bewertung mit dem Quadas-Tool ergab, darf man nicht ausser Acht lassen, dass das Biasrisiko von Observationsstudien gegenüber höherwertigen methodischen Designs wie beispielsweise RCT- Studien eher erhöht ist. Demzufolge muss die eben erwähnte Aussage also mit Vorsicht genossen werden (Richter, 2006).

Als eine weitere Stärke unserer Review betrachten wir, dass wir einerseits bei der Auswahl der Studien, bei der Datenextraktion aber auch bei der gesamten Verfassung der Review sehr systematisch vorgegangen sind.

Die zahlreich Literatur, die es bezüglich des ISG und seinen Schmerzprovokationstests gibt, hat uns erlaubt, uns ein umfangreiches Wissen darüber anzueignen, welches wir anschliessend in unsere Arbeit integrieren konnten. Dies sehen wir als weiteren positiven Aspekt unserer Arbeit.

4.2. Limitationen unserer Review

Aufgrund mangelnder Homogenität der Outcomes aber auch der Clusters in den eingeschlossenen Studien, war es nicht möglich, die Resultate zu poolen und in einem Forrest Plot darzustellen. Wir haben uns dann gezwungenermassen für eine qualitative Beschreibung der Resultate entschieden. Diese Verschiedenheit der Clusters und der beschriebenen Outcomes erschwert die Interpretation sehr stark und macht eine allgemein gültige Aussage bezüglich der Reliabilität und der Validität der Clusters unmöglich. Nicht nur die Outcomes und die Clusters, sondern auch die Art wie die Schmerzprovokationstests durchgeführt werden, ist in den eingeschlossenen Studien zum Teil verschieden. Diese Abwesenheit einer standardisierten Testdurchführung erschwert zusätzlich den Vergleich der Studien untereinander, was eine allgemein gültige Aussage über die Resultate erneut schwierig macht. Dies sehen wir als grösste Limite unserer Review an.

Wie wir bereits in der Einleitung erwähnt haben, fehlt bis anhin ein Goldstandart für das Erkennen von ISG-Problemen (Stuber, 2007). Dies stellt eine weitere Problematik der Review dar, da aufgrund dessen eine gute Kontrolle der Resultate der Schmerzprovokationsteste nur limitiert möglich ist (Ozgocmen u. a., 2008).

Unsere Suche in den fünf Datenbanken, sowie die Suche ausserhalb der elektronischen Datenbanken limitierten wir bezüglich unserer Sprachkenntnisse auf deutsche, französische und englische Studien. Wir können nicht ausschliessen, dass relevante Studien für unsere Review aufgrund dieser Limitation nicht eingeschlossen worden sind. Durch ein systematisches Vorgehen bei der Suche, sowie durch die Suche in mehreren Datenbanken mit einer grossen Variation an Suchbegriffen, die Suche ausserhalb elektronischer Datenbanken und das Kontaktieren von mehrere Autoren (Laslett ; Szadek; Nick Brown; Peter van der Wurff), um unsere Suche zu vervollständigen, haben wir versucht dieses Risiko so klein wie möglich zu halten.

Schlussendlich stellte sicherlich auch die Tatsache, dass diese Review die erste von uns verfasste Review war, eine gewisse Limitation dar. Obschon wir durch verschiedenste Fachpersonen unterstützt wurden, fehlte uns viel Erfahrung und sicherlich auch viel Fachwissen. Dies empfanden wir besonders was die Statistische Auswertung und Analyse der Resultate betraf als einen limitierenden Faktor. Glücklicherweise konnten wir

diesbezüglich aber, wie bereits erwähnt, auf kompetente Unterstützung zählen. Dennoch hat uns diese Tatsache sicherlich in gewissem Masse eingeschränkt.

4.3. Unterschiede zu bisher durchgeführten Reviews

Es gibt bereits einige zu diesem Thema verfasste Reviews (Cattley, Winyard, Trevaskis, & Eaton, 2002; Stuber, 2007; Szadek, Peter van der Wurff, van Tulder, Zuurmond, & Perez, 2009; P van der Wurff, R H Hagmeijer, & W Meyne, 2000; P. van der Wurff, W. Meyne, & R. H. M. Hagmeijer, 2000), was zeigt, dass die Reliabilität und die Validität von Schmerzprovokationstests schon seit längerer Zeit ein zentrales Thema der wissenschaftlichen Forschung darstellen. Zur kritischen Analyse unserer Review haben wir jene mit der aktuellsten Review bezüglich der Suchstrategie, der methodologischen Qualität und der Resultate verglichen und die wichtigsten Unterschiede nachfolgend erwähnt.

Die Review von Szadek et al., welche die aktuellste Review bezüglich der Validität von Schmerzprovokationstests ist, hat die Suche in den Datenbanken Pubmed, Embase und Cinahl durchgeführt. Die Review beschreibt die Outcomes von fünf Studien, welche die Genauigkeit von Schmerzprovokationstests untersuchten. Drei der fünf Studien stimmen mit den von uns eingeschlossenen Studien überein. Die Suche in unterschiedlichen Datenbanken oder die Tatsache, dass bei Szadek et al. nur Studien eingeschlossen wurden, welche als Referenztest eine Kontrastmittelinjektion verwendeten, könnte diesen Unterschied bezüglich der eingeschlossenen Artikel erklären. Dieser Unterschied der eingeschlossenen Studienzahl kann sich aber auf die Resultate nicht gesamthaft auswirken, da in unserer Review wie auch in der Review von Szadek et al. mit einer Ausnahme ein Pooling nicht sinnvoll war. Die Ausnahme stellt das von Szadek gepoolte Resultat für drei oder mehr positive Tests dar. Die methodologische Qualität der eingeschlossenen Studien haben wir höher bewertet als die Review von Szadek. Dies ist möglicherweise darauf zurückzuführen, dass die Fragen des Quidas-Tools, wie wir in unserer Review bereits erwähnt haben, sehr unterschiedlich interpretiert werden können (Whiting u. a., 2003). Bezüglich der Resultate kann kein direkter Vergleich gemacht werden, da die eingeschlossenen Studien zum Teil verschieden sind und die Resultate sowohl in unserer Review wie auch bei Szadek nicht direkt untereinander verglichen werden können. Szadek beschränkt sich in seiner Review bezüglich der Datenextraktion auf Sensitivität, Spezifität und diagnostische Odds Ratio (DOR). Dennoch können einige gemeinsame Tendenzen und/oder Unterschiede der beiden Reviews bezüglich Sensitivität und Spezifität aufgezeigt werden. Sowohl in unserer Review als auch in der Review von

Szadek beschreibt die Studie von Laslett et al. die höchste Sensitivität mit 1.00 für ein oder mehr positive Tests von folgenden sechs Tests: Distractionstest, Thigh thrust Test, Gaenslen Test (rechts und links), Kompressionstest, Sacral thrust Test. Bezüglich der Spezifität haben wir aufgrund der neu eingeschlossenen Studie von Ozgocmen et al. ein anderes Ergebnis. Szadek beschrieb mit 0.88 die höchste Spezifität in der Studie von Laslett et al 05. Diese Werte ergeben sich für fünf oder mehr und für sechs positive Tests. In unserer Review beschreibt Ozgocmen die höchste Spezifität mit einem Wert von 0.93 für vier von fünf positiven Tests des folgenden Clusters: Gaenslen Test, Patrick-Faber Test, Mennell-Test, Thigh thrust Test und der Sacral thrust Test.

Die von uns verfasste Review ist die erste Review, welche die Resultate bezüglich der Reliabilität und oder/ Validität der Schmerzprovokationstests von Ozgocmen et al. und Arab et al. miteinbezogen hat. Die neueste von uns gefundene Review mit ähnlichem Ziel (Szadek u. a., 2009) hat Studien, die bis im September 2007 erschienen sind, eingeschlossen.

4.4. Interpretation der Resultate

Um die Signifikanz unserer Outcome-Werte zu bestimmen, schauen wir hauptsächlich auf die klinische und nicht auf die statistische Signifikanz. Da der Grenzwert je nach Ermessen des Reviewers festgelegt werden kann, haben wir uns dafür entschieden, dass der untere Wert des 95%-Konfidenzintervalls 0.5 nicht unterschreiten darf. Ansonsten wäre es vergleichbar mit Würfeln und die Testdurchführung könnte weggelassen werden, da die Chance dann 1:1 wäre, dass man richtig beziehungsweise falsch testet. Wir haben die Analyse der Resultate nach den jeweiligen Outcomes aufgeteilt und haben das beste Ergebnis von allen eingeschlossenen Studien erwähnt.

Laslett et al. 2005 haben den Cluster mit der höchsten Sensitivität aller von uns überprüften Studien mit einem Wert von 1. 00 für ≥ 1 positive Test (Mark Laslett u. a., 2005). Da sich das 95%-Konfidenzintervall zwischen 0.81 und 1.00 befindet, hat dieser Cluster eine hohe klinische Signifikanz. Aber man muss beachten, dass die Stichprobe im Verhältnis zu den anderen eingeschlossenen Studien eher mittelmässig ist, was zu einer Verfälschung der Resultate sowie ihrer Interpretation führen könnte.

Die höchste Spezifität haben Ozgocmen et al. 2008 für vier von fünf positiven Tests des Clusters, der Wert beträgt 0.93 (Ozgocmen u. a., 2008). Zudem ist der Cluster signifikant, da sich das 95%-Konfidenzintervall von 0.75 bis 0.98 erstreckt. Die Stichprobe

bei Ozgocmen ist mit 40 Personen die zweitkleinste Stichprobe der von uns eingeschlossenen Studien, was wie oben bereits erwähnt zur Fehlinterpretation führen und weniger aussagekräftig sein kann.

Der höchste positiv prädiktive Wert aller eingeschlossenen Studien wurde bei mehr als drei positiven Tests (links) des Clusters von van der Wurff et al. 2006 gefunden (Peter van der Wurff u. a., 2006). Dieses Cluster ist bezüglich des positiven prädiktiven Werts signifikant, da sich der 95%-Konfidenzintervall zwischen 0.62 und 0.92 befindet. Van der Wurff et al. haben eine relativ grosse Stichprobe und dies führt deswegen zu weniger falschen Interpretationen.

Laslett et al. 2005 haben den höchsten negativen prädiktiven Wert von 1.00 für mehr als ein positiver Test des Clusters. Diese Testkombination weist eine hohe klinische Signifikanz auf, da sich der untere Wert des 95%-Konfidenzintervalls deutlich über 0.5 befindet, er beträgt nämlich 0.79. Wie vorgängig bereits erwähnt, ist die Grösse der Stichprobe dieser Studie mittelmässig im Vergleich zu den Anderen.

Die höchste positive Likelihood Ratio wird von Laslett et al. 2003 bei der selektiven Stichprobe beschrieben und beträgt 6.97 für drei oder mehr positive Tests. Dieser Wert weist eine hohe diagnostische Evidenz auf. (Jaeschke, Guyatt, & Sackett, 1994). Das 95%-Konfidenzintervall beinhaltet sowohl schwache diagnostische Evidenz (2.7) als auch überzeugende diagnostische Evidenz (20.27). Obwohl der Mittelwert der positiven Likelihood Ratio eine hohe diagnostische Evidenz hat, sollte man beachten, dass die Stichprobe mittelmässig gross ist.

Laslett et al. 2005 haben die niedrigste negative Likelihood Ratio im Vergleich mit den von uns eingeschlossenen Studien und sie beträgt 0.00, wenn ein oder mehr Tests positiv sind. Da dieser Wert 0.1 unterschreitet, weist dies auf eine überzeugende diagnostische Evidenz. Das 95%-Konfidenzintervall enthält schwache (0.46) aber auch überzeugende (0.00) diagnostische Evidenz-Werte. Das Resultat ist mit Vorsicht zu geniessen, da die Stichprobe vergleichsweise mittelmässig gross ist.

Bezüglich der Intertesterreliabilität haben Arab et al. den höchsten Kappa-Koeffizient mit einem Wert von 0.88 für drei von drei positiven Tests (rechts). Der höchste PABAK-Koeffizient wird von Kokmeyer et al beschrieben. Er beträgt 0.97 für fünf von fünf positiven Tests. Da bei einem Wert von 1.0 die Ergebnisse zu 100% übereinstimmen, kann man bei beiden Werten von einer guten Intertesterreliabilität sprechen.

Betrachtet man die Ergebnisse der Reviews während der letzten Jahre, sind kontroverse Meinungen vorhanden. Die Review von Van der Wurff et al 2000 äussert sich sehr kritisch bezüglich der Reliabilität und Validität von einzelnen Schmerzprovokationstests. Sieben Jahre später beschreibt Stuber in seiner Review jedoch fünf Tests mit einer Sensitivität und Spezifität von über 60%. Dies sind der Distractionstest, Kompressionstest, Thigh thrust Test, Sacral thrust Test und der Resisted abduction Test. Da die Praxis sich zu der Zeit immer mehr in Richtung der Clusters orientiert, (Stuber et al.) schlägt er jene fünf Tests vor, welche in Zukunft in einem Cluster untersucht werden sollten. In der aktuellsten Review beschreiben Szadek et al. 2009 den Kompressionstest und den Thigh thrust Test als die nützlichsten Tests zur Diagnosestellung einer ISG-Pathologie. Bezüglich der Clusters beschreibt jene Review eine gute Validität für drei oder mehr positive Tests. Die Forschung hat sich folglich während der letzten Jahre vermehrt der Untersuchung von Testkombinationen gewidmet. Die neusten Resultate diesbezüglich haben wir bereits erwähnt.

4.5. Relevanz für den Praktiker

Das Prinzip der Diagnostik aus der evidenz-basierten medizinischen Sicht besteht aus der Durchführung einer Serie von einzelnen Tests. Die einzelnen Tests sind charakterisiert durch die Werte der Sensitivität und der Spezifität. Wenn zusätzlich noch der Prävalenz-Wert bekannt ist, kann man die Prädiktion berechnen. Der Prädiktionswert ermöglicht eine quantitative Prognosestellung.

Das Outcome muss anhand der reliablen Assessmentinstrumente objektiv messbar sein. Die empirische Entscheidungsanalyse ist in der Physiotherapie bereits weit verbreitet. Wenn diese durch objektive Entscheidungsanalyse ersetzt wird, sollte die Physiotherapie eine höhere Evidenzstufe erreichen, sowohl innerhalb der Medizin, als auch gegenüber den Patienten und der Gesellschaft (Seichert & Erhart, 2003).

Wenn ein Therapeut einen Test durchführt, ist es wichtig, dass er weiss, wie er das Outcome interpretieren kann, bzw. soll. Es ist nicht zwingend, wenn ein Test positiv ist, dass der Patient dann auch tatsächlich ein Problem hat (E Cottant, 2004).

Die klinischen Tests sind nicht immer sensitiv und spezifisch genug für eine endgültige Differenzierung von Hypothesen (Antje Hüter-Becker, 2005).

Damit keine symptomatischen Patienten durch falsch negative Testresultate ausgeschlossen werden, werden zuerst hochsensitive Tests gemacht und erst anschliessend hochspezifische Tests, um die Gesunden von den positiv getesteten Kranken zu extrahieren. Das begründet sich

folgendermassen: Der Praktiker generiert nach der Anamnese einige Hypothesen. Wenn er die Hypothese der ISG-Problematik verwerfen will, benutzt er zuerst einen hochsensitiven Test respektive Cluster. Wenn der Test negativ ausfällt, hat der Patient mit grösster Wahrscheinlichkeit keine ISG-Problematik und der Praktiker kann weitere Hypothesen untersuchen. Ist jedoch der Test positiv, könnte es sein, dass der Patient doch eine ISG-Pathologie hat, jedoch kann man nicht ohne weitere Tests diese Hypothese bestätigen. Um eben diese zu beweisen oder zu verwerfen, ist es zu empfehlen, dass der Praktiker einen nächsten und diesmal einen hochspezifischen Test verwendet. Bei einem positiven Resultat ist die Wahrscheinlichkeit sehr gross, dass der Patient eine ISG-Problematik hat, dann war der vorgängige Test falsch positiv. Wenn dieser hochspezifische Test negativ ist, kann man die Hypothese einer ISG-Pathologie nicht ausschliessen.

Die Aussagekraft der Tests hängt sowohl von der Spezifität als auch von der Sensitivität ab. Diese Werte dürfen nicht weniger als 40-50% betragen, damit der oben genannte Ablauf wirklich aussagekräftig ist (Pewsnier u. a., 2001).

In der Physiotherapie werden Tests durchgeführt, um Gesunde von Kranken zu differenzieren. Diese Fähigkeit besitzt die Sensitivität. Umgekehrt funktioniert die Spezifität, sie unterscheidet Kranke von Gesunden. Aufgrund dieser beiden Eigenschaften, sind diese zwei Werte am relevantesten für die Praxis. Eine ausführliche Anamnese ist unerlässlich für eine reliable und valide Diagnosestellung.

4.6. Bedeutung für die Forschung

Während der Verfassung unserer Review ist uns aufgefallen, dass es noch einige Mängel und Probleme gibt, was die Forschung über Schmerzprovokationstests des ISG betrifft. Wie wir bereits mehrfach erwähnt haben, gibt es keinen zuverlässigen Goldstandard. Obwohl Infiltrationen und MRI frequentiert benutzt werden, können falsche Testresultate nicht ausgeschlossen werden. Daher dienen diese Referenztests nur begrenzt zur Differenzialdiagnostik (Beyerlein & Bessler, 2006; Kohl, 2007; Mark Laslett u. a., 2005; Ozgocmen u. a., 2008). Des Weiteren fehlt bis anhin eine standardisierte Vorlage bezüglich der Art der Testdurchführung. Dies stellt eine weitere mögliche Fehlerquelle dar und kann zu unterschiedlichen Resultaten innerhalb der einzelnen Studien führen (Mark Laslett u. a., 2003).

Ein weiteres Diskussionsthema stellt die Grösse der Stichproben dar. Die Stichprobengrösse in den von uns eingeschlossenen Studien reicht von 25 bis 78 Teilnehmern. Neben der Repräsentativität beeinflusst der Umfang die Güte der Stichprobe und damit de-

ren Reliabilität. Je grösser die Stichprobe ist, desto genauer entsprechen die Resultate der Wirklichkeit (Mark Laslett u. a., 2003; Sachs & Hedderich, 2006).

Für künftige Studien erachten wir es als wichtig, die oben genannten Punkte zu berücksichtigen und zu verbessern, um die Aussagekraft der Resultate zu erhöhen. Des Weiteren ergibt sich die Frage, inwiefern die Reihenfolge der Tests innerhalb der Clusters einen Einfluss auf die Reliabilität und die Validität hat. Dies gilt es in künftigen Studien zu untersuchen.

5. Schlussfolgerung

Bezüglich der Resultate dieser Review kann gesamthaft keine Testkombination den anderen übergeordnet werden. Will man aber eine valide diagnostische Entscheidung treffen, eignet sich aufgrund der Resultate des Clusters die von Laslett et al. 2003 am besten, welches folgende Schmerzprovokationstests enthält: Distractionstest, Thigh thrust Test, Gaenslen Test, Kompressionstest und Sacral thrust Test. Sind mindestens drei der fünf Tests positiv, kann man sowohl von einer guten Sensitivität (0.91) als auch von einer guten Spezifität (0.87) sprechen. Die Aussagekraft der Testkombinationen ist aber aufgrund mangelnder Forschungsarbeit, sowie einem fehlenden Goldstandard nicht gross genug, um die Referenztests zu ersetzen. Weitere Forschungsarbeiten mit standardisierter Testdurchführung, sowie mit besserer Vergleichbarkeit der Testkombinationen und der Resultate müssen durchgeführt werden.

6. Bibliographie

- Arab, A. M., Abdollahi, I., Joghataei, M. T., Golafshani, Z., & Kazemnejad, A. (2009). Inter- and intra-examiner reliability of single and composites of selected motion palpation and pain provocation tests for sacroiliac joint. *Manual Therapy*, 14(2), 213-221. doi:10.1016/j.math.2008.02.004
- van den Berg, F. (2003). *Angewandte Physiologie: Schmerzen verstehen und beeinflussen* (2. Aufl., Bd. 1-6, Bd. 4). Georg Thieme Verlag.
- Berthelot, J., Labat, J., Le Goff, B., Gouin, F., & Maugars, Y. (2006a). Provocative sacroiliac joint maneuvers and sacroiliac joint block are unreliable for diagnosing sacroiliac joint pain. *Joint, Bone, Spine: Revue Du Rhumatisme*, 73(1), 17-23. doi:10.1016/j.jbspin.2004.08.003
- Berthelot, J., Labat, J., Le Goff, B., Gouin, F., & Maugars, Y. (2006b). Provocative sacroiliac joint maneuvers and sacroiliac joint block are unreliable for diagnosing sacroiliac joint pain. *Joint, Bone, Spine: Revue Du Rhumatisme*, 73(1), 17-23. doi:10.1016/j.jbspin.2004.08.003
- Beyerlein, & Bessler. (2006). Untersuchung und Therapie des Iliosakralgelenks
Der sichere Weg zur Diagnose, (physiopraxis 6/06), 20 - 24.
- Cattley, P., Winyard, J., Trevaskis, J., & Eaton, S. (2002). Validity and reliability of clinical tests for the sacroiliac joint. A review of literature. *Australasian Chiropractic & Osteopathy: Journal of the Chiropractic & Osteopathic College of Australasia*, 10(2), 73-80.
- Cottant, E. (2004). Trente tests de provocation de la douleur de l'articulation sacro-iliaque. *kinésithérapie, les annales*, (31), 13-38.
- Cottant, E. (2004). La sacro-iliaque : une articulation clé. *kinésithérapie, les annales*, (numéro 31), 14 - 19.
- Demir, M., Mavi, A., Gümüşburun, E., Bayram, M., Gürsoy, S., & Nishio, H. (2007). Anatomical variations with joint space measurements on CT. *The Kobe Journal of Medical Sciences*, 53(5), 209-217.
- Glas, A. S., Lijmer, J. G., Prins, M. H., Bonsel, G. J., & Bossuyt, P. M. M. (2003). The diagnostic odds ratio: a single indicator of test performance. *Journal of Clinical Epidemiology*, 56(11), 1129-1135.
- Goff, L. M., Jeffcott, L. B., Jasiewicz, J., & McGowan, C. M. (2008). Structural and biomechanical aspects of equine sacroiliac joint function and their relationship to clinical disease. *Veterinary Journal (London, England: 1997)*, 176(3), 281-

293. doi:10.1016/j.tvjl.2007.03.005

- Hancock, M. J., Maher, C. G., Latimer, J., Spindler, M. F., McAuley, J. H., Laslett, M., & Bogduk, N. (2007). Systematic review of tests to identify the disc, SIJ or facet joint as the source of low back pain. *European Spine Journal*, 16(10), 1539-1550. doi:10.1007/s00586-007-0391-1
- Hildebrandt, J., Müller, G., & Pfingsten, M. (2005). *Lendenwirbelsäule*. Elsevier, München, Urban & Fischer.
- Hüter-Becker, A., & Dölken, M. (2005). *Untersuchen in der Physiotherapie*. Georg Thieme.
- Hüter-Becker, A. (2005). *Lehrbuch zum neuen Denkmodell der Physiotherapie 2. Bewegungsentwicklung, Bewegungskontrolle*. Georg Thieme Verlag.
- Jaeschke, R., Guyatt, G. H., & Sackett, D. L. (1994). Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA: The Journal of the American Medical Association*, 271(9), 703-707.
- Kane, R., Atherly, A., Schüssler-Fiorenza, C., Constantine, M., Frytak, J., Herbert, P., Holtzmann, J., u. a. (2006). *Understanding Health Care Outcomes Research* (2. Aufl.). Jones & Barlett.
- Kapandji. (2006). *Funktionelle Anatomie der Gelenke* (4. Aufl.). Georg Thieme Verlag.
- Kohl, J. (2007). Das ISG als Schmerzquelle
ermitteln, (9/07).
- Kokmeyer, D. J., Van der Wurff, P., Aufdemkampe, G., & Fickenscher, T. C. M. (2002). The reliability of multitest regimens with sacroiliac pain provocation tests. *Journal of Manipulative and Physiological Therapeutics*, 25(1), 42-48.
- Koschack, J. (2008). Standardabweichung und Standardfehler: der kleine, aber feine Unterschied. *DMW - Deutsche Medizinische Wochenschrift*, 125(Online-Publikation), 912-917. doi:10.1055/s-2008-1073146
- Laslett, M., McDonald, B., Tropp, H., Aprill, C., & Öberg, B. (2006). Übereinstimmung zwischen Diagnosen, die aufgrund klinischer Untersuchungen und gemäß verfügbarer Referenzstandards gewonnen wurden. *manuelletherapie*, 10(3), 140-151. doi:10.1055/s-2006-926918
- Laslett, M. (2008). Evidence-Based Diagnosis and Treatment of the Painful Sacroiliac Joint. *The Journal of Manual & Manipulative Therapy*, 16(3), 142-152.

- Laslett, M., Aprill, C. N., McDonald, B., & Young, S. B. (2005). Diagnosis of sacroiliac joint pain: validity of individual provocation tests and composites of tests. *Manual Therapy*, 10(3), 207-218. doi:10.1016/j.math.2005.01.003
- Laslett, M., Young, S. B., Aprill, C. N., & McDonald, B. (2003). Diagnosing painful sacroiliac joints: A validity study of a McKenzie evaluation and sacroiliac provocation tests. *The Australian Journal of Physiotherapy*, 49(2), 89-97.
- Merskey, H. (1979). Pain terms: a list with definitions and notes on usage. Recommended by the IASP Subcommittee on Taxonomy. *Pain*, 6(3), 249.
- Oesch, P., Hilfiker, R., Keller, S., Kool, J., Schädler, S., Tal-Akabi, A., Verra, M., u. a. (2007). *Assessments in der muskuloskelettalen Rehabilitation*. Hans Huber.
- Ozgocmen, S., Bozgeyik, Z., Kalcik, M., & Yildirim, A. (2008). The value of sacroiliac pain provocation tests in early active sacroiliitis. *Clinical Rheumatology*, 27(10), 1275-1282. doi:10.1007/s10067-008-0907-z
- Pewsner, Bleuer, Bucher, Battaglia, Jüni, & Egger. (2001). Der Intuition auf der Spur? Das Bayes'sche Theorem und die Diagnostik in der Grundversorgung Teil I, (3), 41 - 45.
- Richter, B. (2006). Stärken und Limitationen von randomisierten klinischen Untersuchungen (RCT). *Der Diabetologe*, 2(0), S15-S20. doi:10.1007/s11428-006-0086-y
- Robinson, H. S., Brox, J. I., Robinson, R., Bjelland, E., Solem, S., & Telje, T. (2007). The reliability of selected motion- and pain provocation tests for the sacroiliac joint. *Manual Therapy*, 12(1), 72-79. doi:10.1016/j.math.2005.09.004
- Sachs, L., & Hedderich, J. (2006). *Angewandte Statistik*. Springer.
- Seichert, N., & Erhart, P. (2003). Die „evidence-based“-Entscheidungsanalyse in der Physikalischen und Rehabilitativen Medizin (Teil 1): Sensitivität, Spezifität und Prädiktion. *Thieme e-Journals*, 115-120. doi:10.1055/s-2003-38544
- Stuber, K. J. (2007). Specificity, sensitivity, and predictive values of clinical tests of the sacroiliac joint: a systematic review of the literature, 51(1), 30-41.
- Szadek, K. M., van der Wurff, P., van Tulder, M. W., Zuurmond, W. W., & Perez, R. S. G. M. (2009). Diagnostic validity of criteria for sacroiliac joint pain: a systematic review. *The Journal of Pain: Official Journal of the American Pain Society*, 10(4), 354-368. doi:10.1016/j.jpain.2008.09.014
- Trudelle, P., Gedda, M., & Gouilly, P. (2004). [Physiotherapy in motion]. *Presse Médicale (Paris, France: 1983)*, 33 Spec No, Sp75-76.

- v.Mühlendahl, K. (1998). Odds Ratio (OR) und Relatives Risiko (RR), (3 (3)), 1.
- Whiting, P., Rutjes, A. W., Reitsma, J. B., Bossuyt, P. M., & Kleijnen, J. (2003). The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Medical Research Methodology*, 3, 25-25. doi:10.1186/1471-2288-3-25
- Whiting, P. F., Weswood, M. E., Rutjes, A. W., Reitsma, J. B., Bossuyt, P. N., & Kleijnen, J. (2006). Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. *BMC Medical Research Methodology*, 6, 9-9. doi:10.1186/1471-2288-6-9
- van der Wurff, P., Hagmeijer, R. H., & Meyne, W. (2000). Clinical tests of the sacroiliac joint. A systematic methodological review. Part 1: Reliability. *Manual Therapy*, 5(1), 30-36. doi:10.1054/math.1999.0228
- van der Wurff, P., Meyne, W., & Hagmeijer, R. H. M. (2000). Clinical tests of the sacroiliac joint: A systematic methodological review. Part 2: Validity. *Manual Therapy*, 5(2), 89-96. doi:10.1054/math.1999.0229
- van der Wurff, P., Buijs, E. J., & Groen, G. J. (2006). A multitest regimen of pain provocation tests as an aid to reduce unnecessary minimally invasive sacroiliac joint procedures. *Archives of Physical Medicine and Rehabilitation*, 87(1), 10-14. doi:10.1016/j.apmr.2005.09.023

7. Anhang

7.1. Anhang I: Suchstrategie

7.1.1. Suchbegriffe

7.1.1.1. Cochrane

((reliability) OR (validity) OR (sensitivity) OR (specificity) OR (diagnostic) OR (diagnosis)) AND ((distraction test) OR (compression test) OR (thigh thrust) OR (patrick test) OR (faber test) OR (drop test) OR (gaenslen test) OR (gaenslens test) OR (pelvic torsion) OR (sacral thrust) OR (cranial shear) OR (bilateral internal rotation of the hip joint) OR (passive internal rotation of the hip joint)) AND ((sacroiliac joint) OR (sacro iliac joint))

7.1.1.2. Pubmed

((reliability) OR (validity) OR (sensitivity) OR (specificity) OR (diagnostic) OR ("Sensitivity and Specificity"[Mesh]) OR (diagnos*[Title/Abstract]) OR (diagnosis[MeSH]) OR (diagnostic*[MeSH]) OR (diagnosis, differential[MeSH]) OR (diagnosis[Subheading])) AND ((distraction test) OR (compression test) OR (thigh thrust) OR (patrick test) OR (faber test) OR (drop test) OR (gaenslen test) OR (gaenslen s test) OR (pelvic torsion) OR (sacral thrust) OR (cranial shear) OR (bilateral internal rotation of the hip joint) OR (passive internal rotation of the hip joint)) AND ((sacroiliac joint) OR (sacroiliac joint*) OR (sacroiliac*) OR (sacro iliac*) OR (sacro iliac joint)))

7.1.1.3. Cinahl

((reliability) OR (validity) OR (sensitivit) OR (specificity) OR (diagnostic) OR ("Sensitivity and Specificity"[Mesh]) OR (diagnos*[Title/Abstract]) OR (diagnosis[MeSH:noexp]) OR (diagnostic *[MeSH:noexp]) OR (diagnosis, differential[MeSH:noexp]) OR (diagnosis[Subheading:noexp])) AND ((distraction test) OR (compression test) OR (thigh thrust) OR (patrick test) OR (faber test) OR (drop test) OR (gaenslen test) OR (gaenslens test) OR (pelvic torsion) OR (sacral thrust) OR (cranial shear) OR (bilateral internal rotation of the hip joint) OR (passive internal rotation of the hip joint)) AND ((sacroiliac joint) OR (sacroiliac joint*) OR (sacroiliac*) OR (sacroiliac*) OR (sacro iliac joint)) AND Humans

7.1.1.4. Psychinfo

(sacroiliac joint) AND (diagnosis) AND ((reliability) OR (specificity) OR (sensitivity) OR (validity))

7.1.1.5. Web of Science

(sacroiliac joint) AND (diagnosis) AND ((reliability) OR (specificity) OR (sensitivity) OR (validity))

7.2. Anhang II: Definitionen der Outcomes

	Definition
Intratesterreliabilität	Wenn wiederholte Messungen unter den selben Bedingungen zu gleichen Ergebnissen führen, ist ein Messverfahren reliabel. (Wiederholung durch gleichen Untersucher) → gemessen anhand Korrelationskoeffizienten (wie z.B. Kappa, PABAK, ...) (Oesch u. a., 2007)
Intertesterreliabilität	Wenn wiederholte Messungen unter denselben Bedingungen zu gleichen Ergebnissen führen, ist ein Messverfahren reliabel. (Wiederholung durch zweiten Untersucher) → gemessen anhand Korrelationskoeffizienten (wie z.B. Kappa, PABAK, ...) (Oesch u. a., 2007)
Kappa-Koeffizient	= Korrelationswert, er zeigt die Stärke des Zusammenhangs zweier Testresultate auf. Er befindet sich zwischen 0.00 (keine Korrelation) und 1.00 (perfekte Korrelation) (Oesch u. a., 2007).
PABAK	= prevalence-adjusted and bias-adjusted kappa (Arab u. a., 2009)
Spezifität	= Wahrscheinlichkeit, dass ein kranker Patient ein positives Testergebnis hat: $\text{Anzahl richtig negativer} / (\text{Anzahl richtig negativer} + \text{Anzahl falsch positiver}) \rightarrow \text{Richtig Positive Rate (0-100\% oder 0-1)}$ (Oesch u. a., 2007)
Sensitivität	= Wahrscheinlichkeit, dass ein gesundes Individuum ein negatives Testergebnis aufweist: $\text{Anzahl richtig positiver} / (\text{Anzahl richtig positiver} + \text{Anzahl falsch negativer}) \rightarrow \text{Richtig Negative Rate (0-100\% oder 0-1)}$ (Oesch u. a., 2007)
Positive (LR+) und negative (LR-) Likelihood Ratio	Die Likelihood Ratio ist eine aus Sensitivität und Spezifität kombinierte Grösse und misst den Informationszuwachs durch ein positives/negatives Testergebnis. $\text{LR+} = \text{Sensitivität} / (1 - \text{Spezifität})$, zeigt die Zunahme der Erkrankungswahrscheinlichkeit bei positivem Testresultat (Oesch u. a., 2007). $\text{LR-} = (1 - \text{Sensitivität}) / \text{Spezifität}$, zeigt die Abnahme der Erkrankungswahrscheinlichkeit bei negativem Testresultat (Oesch u. a., 2007). $\text{LR+} > 10$ bzw. $\text{LR-} < 0,1$ „überzeugende diagnostische Evidenz“ $\text{LR+ } 5 - 10$ bzw. $\text{LR- } 0,1 - 0,2$ „hohe diagnostische Evidenz“ $\text{LR+ } 2 - 5$ bzw. $\text{LR- } 0,2 - 0,5$ „schwache diagnostische Evidenz“ $\text{LR+ } 1 - 2$ bzw. $\text{LR- } 0,5 - 1$ „kaum relevante diagnostische Evidenz“ (Jaeschke u. a., 1994)
Positiv prädiktiver Wert	Indikator der Wahrscheinlichkeit, dass jemand wirklich krank ist, wenn der Test positiv ist. $\text{Anzahl richtig positiver} / (\text{Anzahl der richtig positiven} + \text{Anzahl der falsch positiven})$ (Oesch u. a., 2007)
Negativ prädiktiver Wert	Indikator der Wahrscheinlichkeit, dass jemand wirklich gesund ist, wenn der Test negativ ist. $\text{Anzahl richtig negativer} / (\text{Anzahl der richtig negativen} + \text{Anzahl der falsch negativen})$ (Oesch u. a., 2007)
Area under the ROC Curve (AUC)	= Fläche unter der Kurve der Sensitivität in Abhängigkeit zu 1-Spezifität (Kurve des Anteils der richtig Positiven in Abhängigkeit zu den falsch Positiven) (Hancock u. a., 2007). AUC befindet sich zwischen 0.5 und 1. 0.5 ist ein wertloser Wert. 1 ist ein perfekter Wert, d.h. Sensitivität und Spezifität haben den höchstmöglichen Wert (Mark Laslett u. a., 2005).
Odds Ratio	Odds Ratio wird auch Quotenverhältnis oder Chancenverhältnis genannt Odds Ratio ist eine statistische Masszahl. Hierbei wird der Zusammenhang von zwei Merkmalen angeschaut, z.B. werden Personen mit Risikofaktoren verglichen mit Personen ohne Risikofaktoren (Vier-Felder-Tabelle) (v.Mühlendahl, 1998).
Diagnostic Odds Ratio	Diagnostic Odds Ratio ist ein Einzelindikator von der Testleistung. Die Formel lautet: $((\text{Richtig positiv}) * (\text{richtig negativ})) / ((\text{falsch positiv}) * (\text{falsch negativ}))$ (Glas, Lijmer, Prins, Bonsel, & Bossuyt, 2003)
Standardfehler (SE)	= Messungenauigkeit. Er sagt etwas über die Genauigkeit des Mittelwerts in einer Stichprobe aus. Der Standardfehler ist von der Stichprobengrösse abhängig, wenn die Stichprobe verdoppelt wird, halbiert sich der Standard. Je näher der Wert gegen 0 geht, umso kleiner ist die Abweichung (Koschack, 2008).

Tabelle 5: Definitionen der Outcomes

7.3. Anhang III: Testbeschreibungen

7.3.1. Bilaterale Innenrotation der Hüfte

Der Patient befindet sich in Bauchlage. Der Therapeut bringt die Beine in maximale Innenrotation bei 90° Knieflexion und führt einen leichten Überdruck aus, so dass das ISG strapaziert wird (H. S. Robinson u. a., 2007).



(H. S. Robinson u. a., 2007)

7.3.2. Distractionstest = Gapping-Test = test d'Ericksen



(Mark Laslett u. a., 2003)

Der Patient befindet sich in Rückenlage. Der Therapeut übt einen nach posterior gerichteten bilateralen Druck auf die Spina Iliaca Anterior Superior (SIAS) aus. Man nimmt an, dass das ISG anterior gedehnt und posterior komprimiert wird (Mark Laslett u. a., 2003).

Der Therapeut übt den Druck mit gekreuzten Armen aus (Kokmeyer u. a., 2002; Ozgocmen u. a., 2008; H. S. Robinson u. a., 2007; Peter van der Wurff u. a., 2006)

7.3.3. Drop Test

Der Drop Test ist der einzige aktive Schmerzprovokationstest in unserer Studie. Der Patient führt eine standardisierte Bewegung aus, welche den Schmerz verstärken oder reproduzieren kann. Der Patient befindet sich im Einbeinstand, hebt die Ferse vom Boden ab und lässt sich dann wieder auf die angehobene Ferse zurückfallen (H. S. Robinson u. a., 2007).



(H. S. Robinson u. a., 2007)

7.3.4. Gaenslen Test = Pelvic torsion = Mennell-Test in Rückenlage

Der Patient befindet sich an der Tischkante in Rückenlage. Ein Bein hängt über den Tischrand, das andere ist in Knie und Hüft gebeugt. Die Hände des Therapeuten befinden sich auf dem rechten beziehungsweise linken Knie und üben einen Druck in Richtung



(Mark Laslett u. a., 2003)

der Brust des Patienten (gebeugtes Bein) beziehungsweise Richtung Boden (gestrecktes Bein) aus. Der Test wird beidseitig durchgeführt. Er erzeugt eine posteriore Rotationskraft des ISG auf der flektierten Seite und eine anteriore Rotationskraft auf der gestreckten Seite (Mark Laslett u. a., 2003).

Beide Knie des Patienten werden gebeugt, das kontralaterale Bein wird in maximaler Flexion und leichter Abduktion fixiert. Anschliessend bringt der Therapeut das ipsilaterale Bein in Extension, indem er einen leichten Druck in Richtung der Extension gibt (Kokmeyer u. a., 2002; Peter van der Wurff u. a., 2006).

7.3.5. Kompressionstest = Volkman-Test = Approximation test

Der Patient befindet sich in Seitenlage, Knie und Hüft befinden sich in etwa 90° Flexion. Der Therapeut führt einen vertikal gerichteten Druck auf die oben liegende Crista iliaca aus. Der Test komprimiert das ISG (Mark Laslett u. a., 2003).



(Mark Laslett u. a., 2003)

Hüft in 45° Flexion (Kokmeyer u. a., 2002; Ozgocmen u. a., 2008; Peter van der Wurff u. a., 2006).



(H. S. Robinson u. a., 2007)

Der Therapeut übt Druck auf das Becken aus, indem er sich mit dem Oberkörper über das Becken des Patienten legt und mit seiner Brust das Sakrum komprimiert (H. S. Robinson u. a., 2007)

7.3.6. Mennell-Test

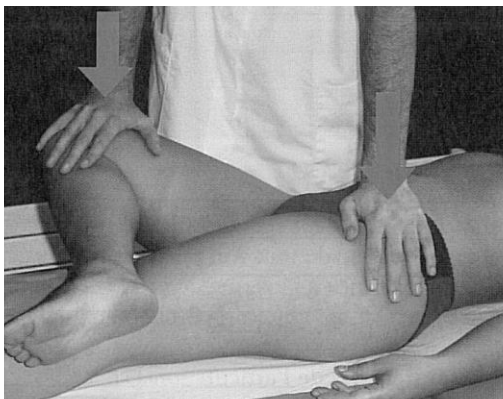
Der Patient befindet sich in Seitenlage, die zu testende Seite befindet sich unten. Das unten liegende Bein wird in Hüft und Knie gebeugt. Der Therapeut befindet sich hinter dem Patienten. Eine Hand befindet sich auf der ipsilateralen Glutealregion sowie auf der Crista Iliaca. Die andere Hand befindet sich auf dem ipsilateralen Knie und übt einen leicht forcierten Druck in die Extension des Beines aus (Ozgocmen u. a., 2008).



(Liebau, 2010)

7.3.7. Patrick-Faber Test

Der Patient befindet sich in Rückenlage. Der Patient bringt die Hüfte der zu testenden Seite in Flexion, Abduktion und Aussenrotation und das Knie in Flexion, so dass die Ferse auf dem kontralateralen Knie zu liegen kommt. Anschliessend fixiert der Therapeut die kontralaterale Spina iliaca anterior superior und übt einen nach unten gerichteten Druck auf das gebeugte Knie aus (Arab u. a., 2009; Ozgocmen u. a., 2008; H. S. Robinson u. a., 2007).



(E Cottant, 2004)

Mediale Fersenseite des ipsilateralen Beines wird an das gestreckte Knie gelegt (Kokmeyer u. a., 2002; Peter van der Wurff u. a., 2006).

7.3.8. Resisted abduction Test



(Liebau, 2010)

Der Patient befindet sich in Rückenlage. Das Bein auf der zu testenden Seite befindet sich in Extensionsnullstellung sowie in 30° Abduktion. Der Therapeut übt einen Druck in Richtung Adduktion aus, welchem der Patient entgegenhalten muss (Arab u. a., 2009).

7.3.9. Sacral thrust Test



(Laslett u. a., 2003)

Der Patient befindet sich in Bauchlage. Der Therapeut übt einen nach ventral gerichteten Druck auf das Zentrum des Sakrums aus. Es wird ein von posterior nach anteriore Scherkraft in Beiden Iliosakralgelenken erzeugt (Mark Laslett u. a., 2003).

Beine in leichter Abduktionsstellung (Ozgocmen u. a., 2008).

7.3.10. Thigh thrust Test =Posterior shear=P4-Test (posterior pelvic pain provocation test)

Der Patient befindet sich in Rückenlage. Der Therapeut flektiert Hüft und Knie, so dass sich die Hüfte in 90° Flexion befindet und bringt das Bein in eine leichte Adduktion. Mit seiner Hand und seinem Arm umfasst er das gebeugte Knie. Mit der anderen Hand umfasst er das Sakrum. Diese Position beibehaltend, übt er einen Druck in Richtung der Längsachse des Femurs aus. Dies führt zu einer von anterior nach posterior gerichteten Scherkraft im Sakrum (Arab u. a., 2009; Kokmeyer u. a., 2002; Mark Laslett u. a., 2003; Peter van der Wurff u. a., 2006).



(Laslett u. a., 2003)

Das Bein ist nicht adduziert (Ozgocmen u. a., 2008).

7.3.11. Unilaterale Innenrotation der Hüfte



(Liebau, 2010)

Der Patient befindet sich in Bauchlage. Der Therapeut bringt das Bein in maximale Innenrotation bei 90° Knieflexion und führt einen leichten Überdruck aus, so dass das ISG strapaziert wird (H. S. Robinson u. a., 2007).

7.4. Anhang IV: Quadas Tool

Item	Yes	No	Unclear
1. Was the spectrum of patients representative of the patients who will receive the test in practice?	()	()	()
2. Were selection criteria clearly described?	()	()	()
3. Is the reference standard likely to correctly classify the target condition?	()	()	()
4. Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?	()	()	()
5. Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?	()	()	()
6. Did patients receive the same reference standard regardless of the index test result?	()	()	()
7. Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?	()	()	()
8. Was the execution of the index test described in sufficient detail to permit replication of the test?	()	()	()
9. Was the execution of the reference standard described in sufficient detail to permit its replication?	()	()	()
10. Were the index test results interpreted without knowledge of the results of the reference standard?	()	()	()
11. Were the reference standard results interpreted without knowledge of the results of the index test?	()	()	()
12. Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?	()	()	()
13. Were uninterpretable/ intermediate test results reported?	()	()	()
14. Were withdrawals from the study explained?	()	()	()

(Whiting u. a., 2003)

7.5. Anhang V: Tabellen der Einzeltests aus Studien

Tests	Side	Tester 1					Tester 2					Inter-tester				
		κ (SE)	95% CI	κ_{\max}	PI (BI)	PABAK	κ (SE)	95% CI	κ_{\max}	PI (BI)	PABAK	κ (SE)	95% CI	κ_{\max}	PI (BI)	PABAK
Gillet	R	0.42 (0.22)	−0.01 to 0.87	0.44	0.56 (0.12)	0.6	0.25 (0.26)	−0.2 to 0.77	0.25	0.6 (0)	0.52	0.41 (23)	0.03–0.87	0.88	0.56 (0.04)	0.6
	L	0.49 (0.2)	0.09–0.89	0.70	0.48 (0.12)	0.6	0.23 (0.22)	−0.2 to 0.67	0.61	0.44 (0.16)	0.36	0.34 (0.21)	−0.06 to 0.7	0.72	0.4 (0.12)	0.44
Standing flexion	R	0.68 (0.16)	0.35–1.01	0.89	0.48 (0.04)	0.76	0.6 (0.21)	0.18–1.02	0.6	0.64 (0.12)	0.76	0.51 (0.22)	0.08–0.95	0.51	0.6 (0.16)	0.68
	L	0.61 (0.17)	0.27–0.96	0.75	0.44 (0.16)	0.68	0.51 (0.22)	0.08–0.95	0.51	0.6 (0.16)	0.68	0.55 (0.17)	0.2–0.9	0.73	0.32 (0.04)	0.6
Sitting flexion	R	0.73 (0.14)	0.45–1.01	0.91	0.32 (0.04)	0.76	0.65 (0.18)	0.29–1.02	0.65	0.56 (0.12)	0.76	0.75 (0.16)	0.42–1.08	0.75	0.6 (0.08)	0.84
	L	0.65 (0.15)	0.34–0.96	0.82	0.28 (0.08)	0.68	0.56 (0.17)	0.21–0.9	0.73	0.32 (0.12)	0.6	0.64 (0.16)	0.32–0.96	0.64	0.36 (0.16)	0.68
Prone knee flexion	R	0.41 (0.18)	0.07–0.78	0.75	0.24 (0.12)	0.44	0.34 (0.21)	−0.06 to 0.7	0.72	0.4 (0.12)	0.44	0.58 (0.16)	0.25–0.91	0.75	0.24 (0.12)	0.6
	L	0.27 (0.25)	−0.22 to 0.78	0.52	0.6 (0.16)	0.52	0.48 (0.2)	0.07–0.88	0.89	0.48 (0.04)	0.6	0.33 (0.26)	−0.18 to 0.85	0.61	0.64 (0.12)	0.6
FABER	R	0.41 (0.18)	0.07–0.78	0.75	0.24 (0.12)	0.44	0.31 (0.2)	−0.08 to 0.70	0.74	0.28 (0.08)	0.36	0.44 (0.19)	0.06–0.83	1	0.36 (0)	0.52
	L	0.40 (0.19)	0.03–0.78	0.91	0.24 (0.04)	0.44	0.4 (0.21)	−0.03 to 0.82	0.80	0.44 (0.08)	0.52	0.49 (0.2)	0.09–0.89	0.70	0.48 (0.12)	0.6
Thigh thrust	R	0.44 (0.19)	0.06–0.83	1	0.36 (0)	0.52	0.49 (0.2)	0.09–0.89	0.62	0.48 (0.12)	0.6	0.6 (0.18)	0.24–0.96	1	0.44 (0)	0.68
	L	0.4 (0.21)	0.00–0.82	0.80	0.44 (0.08)	0.52	0.51 (0.22)	0.08–0.95	0.51	0.6 (0.16)	0.68	0.4 (0.21)	0.00–0.82	0.80	0.44 (0.08)	0.52
Resisted abduction	R	0.48 (0.2)	0.07–0.88	0.79	0.48 (0.04)	0.6	0.56 (0.19)	0.17–0.95	0.78	0.52 (0.08)	0.68	0.78 (0.14)	0.49–1.07	0.78	0.52 (0.08)	0.84
	L	0.5 (0.22)	0.06–0.95	0.75	0.6 (0.08)	0.68	0.62 (0.25)	0.11–1.12	1	0.76 (0)	0.84	0.5 (0.26)	−0.02 to 1.03	0.84	0.72 (0.04)	0.76

Tabelle 6: Arab et al. 2008

	No. of subjects	MTS threshold	% Agreement	Kappa	95% CI of Kappa	SE	BAK	BI	PABAK	PI
Total	78	3/5	93.59	0.70	0.45-0.95	0.13	0.70	0.01	0.87	-0.76
-asym	59	3/5	91.53	0.69	0.42-0.95	0.13	0.69	0.02	0.83	-0.68
Total	78	1/5	83.33	0.63	0.47-0.83	.093	0.07	-0.09	0.65	-0.36
-asym	59	1/5	81.36	0.64	0.44-0.84	0.10	0.64	-0.08	0.64	-0.25
Total	78	2/5	92.31	0.74	0.54-0.94	0.10	0.74	-0.05	0.85	-0.64
-asym	59	2/5	89.83	0.72	0.51-0.93	0.11	0.72	-0.07	0.80	-0.53
Total	78	4/5	96.15	0.71	0.38-1.03	0.17	0.71	0.01	0.92	-0.86
-asym	59	4/5	94.92	0.70	0.37-1.03	0.17	0.70	0.02	0.90	-0.81
Total	78	5/5	98.72	0.66	0.00-1.32	0.34	0.66	-0.01	0.97	-0.96
-asym	59	5/5	98.31	0.66	0.00-1.32	0.34	0.66	-0.02	0.97	-0.95
Distraction test	78	NA	88.46	0.46	0.13-0.79	0.17	0.46	0.01	0.46	-0.76
-asym	59	NA	84.75	0.45	0.10-0.78	0.17	0.45	0.02	0.44	-0.68
Compression test	78	NA	93.59	0.58	0.23-0.94	0.18	0.58	0.01	0.87	-0.83
-asym	59	NA	91.52	0.57	0.21-0.93	0.19	0.57	0.02	0.83	-0.78
Gaenslen test	78	NA	89.74	0.58	0.32-0.88	0.14	0.58	-0.12	0.76	-0.72
-asym	59	NA	88.14	0.60	0.33-0.88	0.14	0.59	-0.12	0.76	-0.64
Patrick sign	78	NA	92.31	0.62	0.33-0.91	0.15	0.62	-0.03	0.85	-0.76
-asym	59	NA	89.83	0.61	0.31-0.91	0.15	0.61	-0.03	0.80	-0.69
Thigh thrust	78	NA	88.46	0.69	0.51-0.88	0.10	0.69	-0.06	0.77	-0.50
-asym	59	NA	86.44	0.67	0.46-0.88	0.11	0.67	-0.07	0.73	-0.42

Tabelle 7: Kokmeyer et al. 2002

Prevalence, sensitivity, specificity and likelihood ratios for individual SIJ provocation tests						
	Distraction	Compression	Thigh thrust	Gaenslen's (right)	Gaenslen's (left)	Sacral thrust
Prevalence of positive test	31.9%	43.8%	50.0	37.0%	31.1%	37.5%
Sensitivity	0.60	0.69	0.88	0.53	0.50	0.63
95% CI	0.36, 0.80	0.44, 0.86	0.64, 0.97	0.30, 0.75	0.27, 0.73	0.39, 0.82
Specificity	0.81	0.69	0.69	0.71	0.77	0.75
95% CI	0.65, 0.91	0.51	0.82	0.53, 0.84	0.60, 0.89	0.58, 0.87
PPV	0.60	0.52	0.58	0.47	0.50	0.56
95% CI	0.36, 0.80	0.32, 0.72	0.39, 0.76	0.26, 0.69	0.27, 0.73	0.34, 0.75
NPV	0.81	0.82	0.92	0.76	0.77	0.80
95% CI	0.65, 0.91	0.63, 0.92	0.74, 0.98	0.58, 0.88	0.60, 0.89	0.63, 0.91
+ LR	3.20	2.20	2.80	1.84	2.21	2.50
95% CI	1.42, 7.31	1.18, 4.09	1.66, 4.98	0.87, 3.74	0.95, 5.00	1.23, 5.09
−LR	0.49	0.46	0.18	0.66	0.65	0.50
95% CI	0.24, 0.83	0.20, 0.87	0.05, 0.55	0.34, 1.09	0.34, 1.03	0.24, 0.87
Notes: PPV = positive predictive value, NPV = negative predictive value, + LR = likelihood ratio for positive test, −LR = likelihood ratio for negative test, 95% CI = 95% confidence interval.						

Tabelle 8: Laslett et al. 2005

	Gaenslen	Patrick-Faber	Mennell	Compression	Thigh thrust	Sacral thrust
Right						
Sensitivity	0.44	0.66	0.66	0.22	0.55	0.33
95% CI	0.15–0.77	0.30–0.90	0.30–0.90	0.03–0.59	0.22–0.84	0.09–0.69
Specificity	0.80	0.51	0.80	0.83	0.70	0.74
95% CI	0.61–0.91	0.33–0.69	0.61–0.91	0.65–0.93	0.51–0.85	0.55–0.87
PPV	0.40	0.28	0.50	0.28	0.35	0.27
95% CI	0.13–0.72	0.12–0.52	0.22–0.77	0.05–0.69	0.13–0.64	0.07–0.60
NPV	0.83	0.84	0.89	0.78	0.84	0.79
95% CI	0.64–0.93	0.59–0.95	0.70–0.97	0.60–0.90	0.64–0.94	0.59–0.91
+LR	2.29	1.37	3.44	1.37	1.91	1.29
95% CI	0.82–6.39	0.76–2.48	1.49–8.09	0.31–5.94	0.85–4.27	0.42–3.88
	0.68	0.64	0.41	0.92	0.62	0.89
95% CI	0.37–1.25	0.24–1.72	0.16–1.05	0.64–1.33	0.29–1.33	0.55–1.45
Left						
Sensitivity	0.36	0.54	0.45	0.27	0.45	0.45
95% CI	0.12–0.68	0.24–0.81	0.18–0.75	0.07–0.60	0.18–0.75	0.18–0.75
Specificity	0.75	0.62	0.86	0.93	0.86	0.89
95% CI	0.56–0.88	0.42–0.78	0.67–0.95	0.75–0.98	0.67–0.95	0.71–0.97
PPV	0.36	0.35	0.55	0.60	0.55	0.62
95% CI	0.12–0.68	0.15–0.61	0.22–0.84	0.17–0.92	0.22–0.84	0.25–0.89
NPV	0.75	0.78	0.80	0.77	0.80	0.81
95% CI	0.56–0.88	0.55–0.91	0.61–0.91	0.59–0.88	0.61–0.91	0.62–0.92
+LR	1.50	1.43	3.29	3.95	3.29	4.39
95% CI	0.54–4.15	0.70–2.93	1.07–10.06	0.76–20.57	1.07–10.06	1.25–15.36
	0.83	0.73	0.63	0.78	0.63	0.60
95% CI	0.52–1.33	0.36–1.45	0.36–1.09	0.54–1.12	0.36–1.09	0.35–1.05

Tabelle 9: Ozgocmen et al. 2008

Inter-examiner agreement on SIJ tests when no pain and discordant pain is merged			
<i>n</i> = 61	% Agreement	Kappa	95% CI
Compression test right side	82	0.48	0.18–0.78
Compression test left side	88	0.67	0.43–0.91
Distraction test	82	0.63	0.43–0.83
P4 right side	84	0.76	0.48–0.86
P4 left side	87	0.74	0.57–0.91
Patrick–Faber test right side	80	0.60	0.39–0.81
Patrick–Faber test left side	74	0.48	0.27–0.69
Bilateral internal rotation	79	0.56	0.33–0.79
Internal rotation right side	90	0.78	0.60–0.94
Internal rotation left side	89	0.88	0.75–1.01
Drop test right side	97	0.84	0.61–1.06
Drop test left side	88	0.47	0.11–0.83

Tabelle 10: Robinson et al. 2005