

**DIE INTERRATER RELIABILITÄT DER SEGMENTALEN PASSIVEN
MOBILITÄTSTESTUNG DER LENDENWIRBELSÄULE IM SITZ MIT
MARKIERTEN PROCESSUS SPINOSI**

Masterthesis zur Erlangung des akademischen Grades

Master of Science

im Universitätslehrgang Osteopathie

eingereicht von

Silvia Maurer

Department für Gesundheitswissenschaften und Biomedizin

Donau-Universität Krems

Betreuerin: **Mag. Dr. Astrid Grant Hay**

Linz, am 23. November 2014

EIDESSTATTLICHE ERKLÄRUNG

Ich, Silvia Maurer, geboren am 20. April 1983 in Freistadt (OÖ) erkläre,

1. dass ich meine Master Thesis selbständig verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt und mich auch sonst keiner unerlaubten Hilfen bedient habe,
2. dass ich meine Master Thesis bisher weder im In-, noch im Ausland in irgendeiner Form als Prüfungsarbeit vorgelegt habe,
3. dass ich, falls die Master Thesis mein Unternehmen oder einen externen Kooperationspartner betrifft, meinen Arbeitgeber über Titel, Form und Inhalt der Master Thesis unterrichtet und sein Einverständnis eingeholt habe.

Datum

Unterschrift

DANKSAGUNG

Mein großer Dank gilt den „Brains“ für die zahlreichen gemeinsamen Lernstunden, die vielen fruchtbaren Diskussionen, die großartige gegenseitige Unterstützung und Motivation in den letzten Jahren, die tolle Mitarbeit bei meiner Masterthese und v.a. für das viele gemeinsame Lachen, welches auch in stressigen Zeiten nie zu kurz kam.

Ich möchte mich bei Fr. Mag. Dr. Astrid Grant Hay für die herausragende Masterthesen-Betreuung bedanken. Ihre große fachliche Kompetenz, ihre ansteckende Begeisterung für wissenschaftliches Arbeiten und ihre herzliche Art machen sie zur besten Betreuerin, die man sich wünschen kann.

Sarah danke ich für ihr großes Verständnis und die wunderbare Unterstützung in den letzten Jahren.

Ich danke der Praxisgemeinschaft Rainerstraße für die Räumlichkeiten, Dr. Gebhart Woissetschläger für die Unterstützung bei der statistischen Auswertung, den Probanden für ihre Teilnahme und den Korrekturleserinnen fürs Korrigieren.

Und last but not least möchte ich mich bei meiner Familie bedanken, weil ich immer auf sie zählen kann.

Aus Gründen der einfacheren Lesbarkeit wird auf die geschlechtsneutrale Differenzierung, z.B. Patient/Patientin, verzichtet. Im Text wird die männliche Form verwendet. Die Angaben beziehen sich jedoch auf Angehörige beider Geschlechter. Sämtliche Fotos wurden mit freundlicher Genehmigung der abgebildeten Personen eingefügt.

INHALTSVERZEICHNIS

ABSTRACT (DT.)	4
1 EINLEITUNG	5
2 FORSCHUNGSSTAND UND THEORIE	7
2.1 RELIABILITÄT UND INTERRATER-RELIABILITÄT	7
2.2 GRUNDLAGEN DER STATISTIK ZUR ERHEBUNG DER INTERRATER-RELIABILITÄT	8
2.3 DIE SEGMENTALE PASSIVE MOBILITÄTSTESTUNG DER LENDENWIRBELSÄULE	9
2.3.1 METHODE	9
2.3.2 BEURTEILUNG	11
2.4 INTERRATER-RELIABILITÄT DER PASSIVEN SEGMENTALEN MOBILITÄTSTESTUNG DER LENDENWIRBELSÄULE	14
2.4.1 REVIEWS	14
2.4.2 QUALITÄTSKRITERIEN	15
2.4.3 INTERRATER-RELIABILITÄT DER SPEZIFISCH SEGMENTALEN PHYSIOLOGISCHEN BEWEGUNGEN DER LENDENWIRBELSÄULE	20
3 FORSCHUNGSFRAGE UND HYPOTHESEN	23
4 METHODIK	25
4.1 LITERATURRECHERCHE UND DATENBANKEN	25
4.2 STUDIENDESIGN	26
4.3 DIE UNTERSUCHER	26
4.4 DIE ASSISTENTINNEN	26
4.5 DIE PROBANDEN	27
4.6 BLINDIERUNG	29
4.7 DIE UNTERSUCHUNG	29
4.7.1 TESTER-ABSTIMMUNG	29
4.7.2 DURCHFÜHRUNGSORT	29
4.7.3 ABLAUF	30
4.7.4 FRAGEBOGEN UND MATERIALIEN	30
4.7.5 MARKIERUNG DER PROCESSUS SPINOSI DER LENDENWIRBELSÄULE	31

4.7.6	DURCHFÜHRUNG DER PASSIVEN SEGMENTALEN PHYSIOLOGISCHEN MOBILITÄTSTESTUNG DER LENDENWIRBELSÄULE	32
4.7.7	FEEDBACK DER TESTER DIREKT NACH DER TESTUNG	34
4.8	DATENAUFBEREITUNG UND ANALYSE	36
4.8.1	DARSTELLUNG DER ERGEBNISSE	36
4.8.2	BERECHNUNG UND BEURTEILUNG VON COHENS KAPPA	36
5	<u>ERGEBNISSE</u>	38
5.1	INTERRATER-RELIABILITÄT IN DER SAGITTALEBENE	38
5.1.1	FLEXION	38
5.1.2	EXTENSION	38
5.1.3	FLEXION UND EXTENSION GESAMT	39
5.1.4	ÜBERSICHT SAGITTALEBENE	39
5.2	INTERRATER-RELIABILITÄT IN DER FRONTALEBENE	40
5.2.1	LATERALFLEXION LINKS	40
5.2.2	LATERALFLEXION RECHTS	40
5.2.3	LATERALFLEXION GESAMT	41
5.2.4	ÜBERSICHT FRONTALEBENE	41
5.3	INTERRATER-RELIABILITÄT IN DER TRANSVERSALEBENE	42
5.3.1	ROTATION LINKS	42
5.3.2	ROTATION RECHTS	42
5.3.3	ROTATION GESAMT	43
5.3.4	ÜBERSICHT TRANSVERSALEBENE	43
5.4	ZUSAMMENFASSENDE DARSTELLUNG DER ERGEBNISSE	44
6	<u>DISKUSSION</u>	46
6.1	VERGLEICH DER ERGEBNISSE MIT DER LITERATUR	46
6.2	LIMITIERENDE FAKTOREN DER METHODOLOGIE	47
6.2.1	STUDIENDESIGN	47
6.2.2	DIE UNTERSUCHER	50
6.2.3	DIE PROBANDEN	50
6.3	AUSBLICK	51
7	<u>KONKLUSION</u>	52

LITERATURVERZEICHNIS	54
TABELLENVERZEICHNIS	61
ABBILDUNGSVERZEICHNIS	61
ABKÜRZUNGSVERZEICHNIS	62
ANHANG	63

**INTERRATER RELIABILITY FOR SEGMENTAL PASSIVE MOTION PALPATION OF THE
LUMBAR SPINE IN SITTING POSITION WITH MARKED SPINOUS PROCESSES**

ABSTRACT (ENGL.)	71
INTRODUCTION	72
METHODS	74
RESULTS	78
DISCUSSION	80
CONCLUSION	83
REFERENCES	84
ANNEX	88

Abstract (dt.)

Die Interrater-Reliabilität der segmentalen passiven Mobilitätstestung der Lendenwirbelsäule im Sitz mit markierten Processus spinosi

Silvia Maurer

Ziel: Die segmentale passive Mobilitätstestung der Wirbelsäule ist eine gängige Untersuchungsmethode, um Restriktionen an der Wirbelsäule aufzuspüren und oft Basis weiterer Therapieentscheidungen. Ziel dieser Studie war es, die Interrater-Reliabilität der spezifisch segmentalen physiologischen Mobilitätstestung der Lendenwirbelsäule im Sitz zu untersuchen.

Methodik: Zwei Untersucher, beide Studenten der Wiener Schule für Osteopathie und erfahrene Physiotherapeuten, testeten die Lendenwirbelsäule von 22 symptomatischen bzw. asymptomatischen Probanden. Vor der Untersuchung wurden die Dornfortsätze von L1 bis S1 von der Studienleiterin markiert. Es wurden die Bewegungsrichtungen Flexion/Extension, Rotation links/rechts und Lateralflexion links/rechts im Sitz untersucht und pro Bewegungsrichtung wurde das hypomobilste Segment bestimmt. Die Probanden waren den Testern völlig unbekannt. Die Blindierung unter den Testern bzw. bezüglich der Symptomatik der Probanden wurde durch zwei unabhängige Assistentinnen gewährleistet, welche auch die Ergebnisse dokumentierten. Die Berechnung der Interrater-Reliabilität erfolgte nach Cohens Kappa für zwei Untersucher in den drei Bewegungsebenen und für die einzelnen Bewegungsrichtungen.

Ergebnisse: Die Interrater-Reliabilität der Mobilitätstestung der Lendenwirbelsäule war schlecht und lag für sämtliche Bewegungsrichtungen und Ebenen im Bereich der Zufallswahrscheinlichkeit. Die Kappa-Werte lagen zwischen $k = 0,04$ und $k = 0,13$. Ein Tester fand in $85 \pm 5\%$ der Fälle eine Bewegungseinschränkung, hingegen der andere nur in $48 \pm 11\%$ der Fälle.

Konklusion: Therapieentscheidungen basierend auf Ergebnissen der passiven physiologischen Mobilitätstestung in der Lendenwirbelsäule sollten kritisch reflektiert werden. Die Übertragbarkeit dieser Ergebnisse auf die Gesamtheit der Osteopathen scheint jedoch fraglich, da die Ergebnisse möglicherweise durch einen Ausreißer beeinträchtigt wurden.

Schlüsselwörter: Mobilitätstestung, Palpation, Lendenwirbelsäule, Interrater-Reliabilität, segmentale Untersuchung

1 EINLEITUNG

Die passive Mobilitätstestung der Lendenwirbelsäule ist eine gängige manuelle Untersuchungsmethode, welche in der Physiotherapie, Osteopathie und anderen manuellen Konzepten gelehrt wird, um Restriktionen zwischen zwei Wirbeln zu finden. Die Ausgangspositionen für diese Untersuchungen reichen in der Praxis vom Sitz über die Seitenlage bis zur Bauchlage. Bisherige Studien zur Reliabilität der Mobilitätstestung wurden oft statisch in Bauchlage unter Ausübung eines posterior-anterior gerichteten Drucks an den Lendenwirbeln getestet (Maher & Adams, 1994; Schneider et al., 2008).

Van Trijffel, Anderegg, Bossuyt, und Lucas (2005) fassten in einer systematischen Review die Interexaminer-Reliabilität der passiven intervertebralen Mobilitätstestung in der zervikalen und lumbalen Wirbelsäule als schlecht bis ausreichend zusammen. Die meisten dieser Studien waren jedoch von schlechter methodologischer Qualität.

Eine mögliche Ursache für die unzureichende Reliabilität könnte laut Snider, Snider, Degenhardt, Johnson, und Kribs (2011) auch darin liegen, dass bereits die schlechte Reliabilität beim Auffinden und Benennen der richtigen Höhe der Processus spinosi die Untersuchungsergebnisse verfälscht. Nur die Markierung der Processus spinosi vor der Untersuchung kann dieser Fehlerquelle entgegenwirken. Das Auffinden und die Zuordnung der Processus spinosi zum richtigen Segment wird durch die Orientierung an bestimmten knöchernen Orientierungspunkten deutlich verbessert (Snider et al., 2011). Nur wenige Studien markierten vor der Durchführung der Untersuchung die Processus spinosi der Lendenwirbelsäule.

Qvistgaard, Rasmussen, Laetgaard, Hecksher-Sorensen, und Bliddal (2007) verwendeten für die Untersuchung der Lendenwirbelsäule Markierungen an L1 und L5, die vom ersten Untersucher angebracht wurden. Die Kappa-Werte für die Interrater-Reliabilität der einzelnen Segmente ergaben dabei für die Extension, Flexion und Lateralflexion eine ausreichende Übereinstimmung. In dieser Studie wurden verschiedene Ausgangsstellungen und Testungen angewendet, um am Ende ein Segment mit somatischer Dysfunktion zu benennen. Deore und May (2012) verwendeten eine Markierung auf den Segmenten L1 bis L5 und führten die Untersuchung statisch mit Posterior-anterior-Druck in Bauchlage durch, mit dem Ergebnis einer schlechten Interrater-Reliabilität.

Auffällig ist, dass kaum eine der gefundenen Studien eine segmentale Untersuchung der Lendenwirbelsäule im Sitz vornimmt - ist dies doch eine günstige Ausgangsposition, um sich einen schnellen Überblick über die Mobilität der Lendenwirbelsäule zu verschaffen. Diese Methode findet sich im „Leitfaden Osteopathie - Parietale Techniken“ (Liem & Dobler, 2010) und wird an

Osteopathischen Schulen in Österreich gelehrt. Auch in der Praxis der Autorin sowie bei vielen Kollegen ist diese immer wieder Teil der Befunderhebung.

Ziel dieser Studie ist es, die Interrater-Reliabilität bei der passiven Mobilitätstestung der Lendenwirbelsäule im Sitz zu untersuchen. Dazu erfolgt eine methodologische Studie, in welcher die Probanden von zwei Osteopathen untersucht werden. Um die Zuverlässigkeit der Untersuchung zu erhöhen, werden die Processus spinosi der Lendenwirbelsäule vorab markiert. Es werden die Bewegungsrichtungen Flexion/Extension, Rotation links/rechts und Lateralflexion links/rechts untersucht.

2 FORSCHUNGSSTAND UND THEORIE

2.1 Reliabilität und Interrater-Reliabilität

In der täglichen Praxis kommt eine Vielzahl an Tests zur osteopathischen, physiotherapeutischen oder manualtherapeutischen Diagnosefindung am Patienten zur Anwendung. Um die Testergebnisse entsprechend gewichten zu können, gibt es verschiedene Gütekriterien.

Anhand von folgenden drei zentralen Gütekriterien lässt sich die Qualität eines Tests beurteilen (Bortz & Döring, 2006; Hussy, Schreier, & Echterhoff, 2009):

- **OBJEKTIVITÄT:** „Bei der **Objektivität** handelt es sich um ein allgemeines Gütekriterium wissenschaftlicher Untersuchungen: Verschiedene Forscher müssen unter den gleichen (Versuchs-) Bedingungen zu den gleichen Ergebnissen gelangen“ (Hussy et al., 2009, S. 22).
- **RELIABILITÄT:** „**Reliabilität** bezeichnet die Zuverlässigkeit und Beständigkeit einer Untersuchung. Reliabel ist ein Instrument dann, wenn es bei einem relativ gleich bleibenden Verhalten gleiche oder ähnliche Ergebnisse liefert“ (Hussy et al., 2009, S. 23).
- **VALIDITÄT:** „Die **Validität** beurteilt eine quantitative Untersuchung danach, ob sie auch gemessen hat, was sie messen wollte“ (Hussy et al., 2009, S. 23).

Die Voraussetzung für Reliabilität ist Objektivität. Ein Test, der nicht objektiv ist, kann auch keine Reliabilität aufweisen. Folglich kann die „Reliabilität [...] maximal so hoch sein wie die Objektivität“ (Bortz & Döring, 2006, S. 196).

Bei der Erhebung der Reliabilität eines Tests unterscheidet man zusätzlich zwischen Intrarater- und Interrater-Reliabilität. **Intrarater-Reliabilität** beschreibt die Höhe der Übereinstimmung, wenn ein Test vom selben Tester wiederholt durchgeführt wird. **Interrater-Reliabilität** gibt die Höhe der Übereinstimmung eines Tests zwischen zwei oder mehreren Untersuchern an [Patijn & International Federation for Manual/Musculoskeletal Medicine (FIMM) Scientific Committee, 2004]. „Die Interrater-Reliabilität ist somit hoch, wenn verschiedene Rater bei den gleichen Test-Personen zu gleichen oder ähnlichen Einschätzungen (Ratings) kommen“ (Hussy et al., 2009, S. 23). In vorliegender Studie wird die Interrater-Reliabilität der manuellen Untersuchung der Lendenwirbelsäule erhoben.

2.2 Grundlagen der Statistik zur Erhebung der Interrater-Reliabilität

Für die statistische Auswertung der Interrater-Reliabilität zwischen zwei Untersuchern empfehlen Patijn und das FIMM Scientific Committee (2004) sowie Bortz und Lienert (2003) die Berechnung des Kappa-Koeffizienten nach Cohen. Kappa-Werte können zwischen $k = -1$ und $k = +1$ liegen, wobei $+1$ die höchste Übereinstimmung beschreibt. Kappa-Werte unter Null entstehen, wenn die Übereinstimmung unter der Zufallserwartung liegt (Bortz & Lienert, 2003).

Bisherige Interrater-Reliabilitätsstudien beschrieben die Übereinstimmung der Tester oft in prozentuellen Angaben. Der Nachteil dabei ist, dass prozentuelle Angaben auch zufällige Übereinstimmungen beinhalten. Dieser Zufall ist höher, je weniger Urteilkategorien verwendet werden (Bortz & Döring, 2006). Auch Sim und Wright (2005) besagen, dass eine Übereinstimmung durch Zufall eigentlich überhaupt keine Übereinstimmung ist. Nur Übereinstimmungen außerhalb der erwarteten, zufälligen Übereinstimmungen entsprechen der „wahren“ Übereinstimmung.

Die Angaben zur Mindest-Stichprobengröße der Teilnehmer sind inhomogen. Laut Bortz & Lienert (2003) genügen für die statistische Berechnung >20 Probanden. Das FIMM scientific committee (2004) hingegen erachtet 40 Probanden als ausreichend.

Bei der Anzahl der Untersucher ist man sich einig, dass zwei ausreichend sind. Für die statistische Erhebung der Interrater-Reliabilität bedarf es nicht mehr. Durch mehrere Untersucher bestünde in der manuellen Untersuchung sogar die Gefahr der Verfälschung der Ergebnisse durch das Eintreten eines therapeutischen Effekts (Patijn & FIMM Scientific Committee, 2004).

2.3 Die segmentale passive Mobilitätstestung der Lendenwirbelsäule

Im folgenden Kapitel werden die unterschiedlichen Methoden zur passiven Untersuchung der Lendenwirbelsäule erklärt und die verschiedenen Beurteilungsparameter erläutert.

2.3.1 Methode

„Die Diagnose ist der wichtigste Teil der osteopathischen Konsultation“ (Ciranna-Raab, Fossum, & Och, 2010, S.70), da eine erfolgreiche Behandlung auf der richtigen Diagnose aufbaut. Die passive segmentale Mobilitätstestung der Wirbelsäule ist ein wichtiges Element der osteopathischen Untersuchung. Dabei werden „die genauen Läsionsparameter einer Dysfunktion [...] qualitativ und quantitativ bewertet, um den Zustand des Gewebes zu beurteilen und die Parameter für die Korrektur aufzustellen“ (Ciranna-Raab et al., 2010, S.72).

Zwei Methoden der segmentalen Untersuchung der Wirbelsäule werden in der Osteopathie oder anderen manualtherapeutischen Konzepten häufig angewendet: spezifische segmentale physiologische Bewegungen und die Testung von Zusatzbewegungen über posterior-anterior gerichteten Druck.

2.3.1.1 Spezifische segmentale physiologische Bewegung (PPIVM)

Die „spezifische segmentale physiologische Bewegung“ oder auch „passive physiologische intervertebrale Bewegung“ (nach Maitland PPIVM = passive physiological intervertebral movements) bewertet die physiologische Beweglichkeit zwischen zwei Wirbeln (Maitland, Hengeveld, & Banks, 2005). Dabei wird der Patient angewiesen zu entspannen, während der Therapeut die Wirbelsäule passiv bewegt und die Mobilität zwischen den Processus spinosi palpiert und beurteilt (Gonnella, Paris, & Kutner, 1982). In der Regel erfolgt diese Untersuchung an der LWS in Seitenlage - der Therapeut bewegt die gebeugten Knie und Hüften des Patienten soweit in Flexion/Extension, bis eine Annäherung oder ein Auseinanderweichen der Processus spinosi tastbar wird. Die Beurteilung erfolgt über das Bewegungsausmaß und die Bewegungsqualität (Ciranna-Raab et al., 2010; Fossum, Sommerfeld, & Vincent, 2010; Wiesner & Bucher-Dollenz, s.a.).

In der osteopathischen Literatur wird diese Testung auch in der Ausgangsposition Sitz beschrieben, wobei hier die Bewegungen der LWS über das Bewegen des Oberkörpers in die verschiedenen Richtungen induziert werden (Fossum et al., 2010).

2.3.1.2 Posterior-anterior gerichteter Druck (PA)

Im manuelltherapeutischen Konzept nach Maitland wird der posterior-anterior gerichtete Druck auch als passive intervertebrale Zusatzbewegung (PAIVM = passive accessory intervertebral movements) bezeichnet (Maitland et al., 2005). Bei der Untersuchung der Zusatzbewegungen der Wirbelsäule befindet sich der Patient in der Regel in Bauchlage. Der Therapeut führt mit der ulnaren Handkante einen von posterior nach anterior gerichteten Druck aus. Dieser kann zentral am Processus spinosus oder unilateral am Processus transversus appliziert werden. Dabei wird das Verhalten von Schmerz, Widerstand und Schutzspasmus in einem Bewegungssegment analysiert. Das Bewegungsausmaß wird im Vergleich zur Norm, zu den Nachbarsegmenten und zur Konstitution des Patienten beurteilt (Wiesner & Bucher-Dollenz, s.a.). Osteopathisch erfolgt die Beurteilung wieder anhand biomechanischer und motorischer Parameter wie Quantität des Gelenkspiels, Qualität der Bewegung und motorischer Reaktionen (Ciranna-Raab et al., 2010).

Eine kritische Betrachtung dieser Untersuchungsmethode erfolgte durch Schomacher (2008). Er untersuchte die Fähigkeit des spezifischen manuellen Bewegens in einzelnen Wirbelsäulensegmenten, wie es bei Maitland gelehrt wird, anhand einer Literaturanalyse. Dabei ging er der Frage nach, ob Physiotherapeuten fähig sind, spezifisch zu bewegen. Sechs Studien, welche die segmentalen Bewegungen durch einen posterior-anterior gerichteten Schub auf den Processus spinosi untersuchten, wurden in die Analyse eingeschlossen. Diese lieferten keine Hinweise darauf, dass PAs spezifisch in einem einzigen Wirbelsäulensegment bewegen, da in den Nachbarsegmenten Mitbewegungen in Form von Flexion und Extension festgestellt wurden. Schomacher stellt somit die Spezifität zur Prüfung der segmentalen Beweglichkeit durch PAs in Frage.

Die meisten Interrater-Reliabilitätsstudien zur segmentalen passiven Mobilitätstestung der LWS untersuchten die sogenannten passiven intervertebralen Zusatzbewegungen in Bauchlage (siehe: Binkley, Stratford, & Gill, 1995; Degenhardt, Johnson, Snider, & Snider, 2010; Degenhardt, Snider, Snider, & Johnson, 2005; Deore & May, 2012; Maher & Adams, 1994; Schneider et al., 2008; Shaw, Dougherty, Treffer, & Glaros, 2012). Degenhardt et al. (2005) führten die PAs an der LWS auch im vorgeneigten Sitz durch.

Huijbregts (2002) fasst in seinem Review zur Mobilitätstestung der LWS auch einige Studien zusammen, welche die PPIVMs in Seitenlage oder im Sitz testeten. Der Großteil der Studien, welche im Sitz durchgeführt wurden, fand vor 1990 statt und ist von schlechter methodischer Qualität.

In zwei weiteren Interrater-Reliabilitätsstudien erfolgte die Kombination von PPIVMs mit PAs zur Identifikation eines einzigen hypomobilen Segments der Lendenwirbelsäule (Arndt, 2013; Qvistgaard et al., 2007).

Zusammengefasst wird ersichtlich, dass sich in jüngster Zeit ein Großteil der Interrater-Reliabilitätsstudien auf die Untersuchung der Zusatzbewegungen konzentriert und weniger auf die physiologischen segmentalen Bewegungen. Da in der Osteopathie die segmentale physiologische Bewegungsuntersuchung jedoch einen hohen Stellenwert in der Auffindung von Dysfunktionen der Wirbelsäule einnimmt, erscheint es sinnvoll, die Reliabilität dieser Methode in qualitativ verbesserten Studien erneut genauer zu untersuchen. Auf die Qualitätskriterien von Reliabilitätsstudien wird im Kapitel 2.4.2 genauer eingegangen. Schomacher (2008) stellt die Untersuchung der Wirbelsäule durch PAs ohnehin kritisch in Frage.

2.3.2 Beurteilung

Die Beurteilung der segmentalen passiven Mobilitätstestung erfolgt anhand biomechanischer und motorischer Parameter (Ciranna-Raab et al., 2010; Schomacher, 2008; Wiesner & Bucher-Dollenz, s.a.):

- **Quantität:** Hierbei wird das Bewegungsausmaß bewertet - üblicherweise mit den Unterteilungen Hypo-, normale und Hypermobilität.
- **Qualität:** Hierbei wird das Endgefühl im Sinne der Elastizität bzw. Flexibilität bewertet.
- **Motorische Reaktionen:** Hierbei wird beurteilt, ob es zu einer frühzeitigen Zunahme der Muskelspannung in einem Segment kommt.

Abbott et al. (2009) erhoben anhand eines Online-Fragebogens in Neuseeland und den USA, was Manual-Physiotherapeuten während der passiven intervertebralen Mobilitätstests zu beurteilen beabsichtigen. Dabei wurde bei der quantitativen Beurteilung zusätzlich zwischen Ausmaß des Bewegungswinkels und Translation des Wirbels unterschieden. Die qualitative Beurteilung unterschied zwischen Widerstand, Endgefühl/Gewebequalität (z.B. Kapsel, Muskel, Bänder...) und Bewegungsqualität während der Bewegung selbst. Weitere Faktoren waren Wirbelposition, Schmerzangaben oder andere.

Die Auswertung ergab, dass bei PPIVMs v.a. Bewegungsqualität während der Bewegung und auch Ausmaß des Bewegungswinkels beurteilt werden, während bei den PAs mehr die Qualität des Widerstandes und die Translation des Wirbels beurteilt werden.

Zusätzlich wurde erhoben, in welchen Kontext die Therapeuten diese Beurteilung stellen. Folgende Möglichkeiten wurden unterschieden:

- Die zu erwartende Mobilität in Bezug auf das Patientenalter bzw. die körperliche Konstitution;
- Die zu erwartende Mobilität verglichen mit anderen Patienten bzw. Gesunden;
- Die zu erwartende Mobilität verglichen mit den Segmenten darüber und darunter;
- Schmerzen des Patienten während der Untersuchung;
- u. a.

Über 60 % der Therapeuten beziehen sich bei der Beurteilung der Mobilität auf das Segment darüber und darunter.

In dieser Studie von Abbott et al. (2009) wird sehr gut ersichtlich, wie viele unterschiedliche Beurteilungsmöglichkeiten zur Verfügung stehen. Diese Breite an Beurteilungsmöglichkeiten findet sich auch in vielen Studien wieder.

Binkley et al. (1995), Deore und May (2012), Gonnella et al. (1982), Landel, Kulig, Fredericson, Li, und Powers (2008), Qvistgaard et al. (2007) und Schneider et al. (2008) beurteilten die Mobilität, jedoch ohne genauere Angaben darüber in welchem Kontext diese steht. Bei den Unterteilungen des Bewegungsausmaßes gibt es auch hier Variationen. In früheren Reliabilitätsstudien wurden noch sehr breit gefächerte Skalen verwendet. Gonella et al. beriefen sich auf die Unterteilungen nach Kaltenborn: 0 (ankylosiert) bis 6 (instabil). Binkley et al. entwickelten eine Skala von 1 (exzessive Mobilität) bis 9 (keine Mobilität) zur Beurteilung der Mobilität, mit der Begründung, dass eine 3-stufige Skala weniger reliabel ist. In späteren Studien wurde die Mobilität von Qvistgaard et al. und Schneider et al. als „normal“ oder „hypomobil“ bewertet, von Deore und May zusätzlich noch als „hypermobil“. Landel et al. (2008) hinterfragen auch die Beurteilung „normal“, da zwei Therapeuten immer unterschiedliche Erfahrungen haben und deshalb auch keine gemeinsame Definition von „normaler Mobilität“ bestehen kann. Die Unterscheidung in „hypomobilstes/hypermobilstes“ Segment wird vorgeschlagen.

Qualitative Beurteilungskriterien wie Widerstand oder Steifigkeit wurden z.B. in den Reliabilitätsstudien von Degenhardt et al. (2010) oder Maher und Adams (1994) herangezogen.

In einem Literatur-Review zur segmentalen Mobilitätstestung von Haneline et al. (2008) wurde zwischen qualitativ versus quantitativ beurteilten Studien unterschieden. Haneline et al. kritisieren vor allem bei qualitativen Erhebungen eine mangelhafte Beschreibung der Methode. Huijbregts (2002) wirft auch das Argument auf, dass es sehr unwahrscheinlich ist, dass zwei unterschiedliche Tester exakt die gleichen Beurteilungsparameter verwenden und in Folge die Interrater-Reliabilität beeinträchtigen.

Die verschiedenen Beurteilungskriterien und zusätzlichen Abstufungen der qualitativen und quantitativen Komponenten spiegeln den Praxisalltag wider. Unterschiedliche manualtherapeu-

tische Konzepte und Beurteilungskriterien erschweren die Vergleichbarkeit der Mobilitätstestung und folglich auch die Vergleichbarkeit von Reliabilitätsstudien zur Mobilitätstestung. Einheitliche Beurteilungen könnten dem Abhilfe schaffen und würden auch die Kommunikation unter Manualtherapeuten erleichtern. Da bei PPIVMs keine direkte translatorische Kraft vom Therapeuten auf den Wirbel aufgebracht, sondern in erster Linie „passiv“ palpiert wird, scheint hier die quantitative Beurteilung am sinnvollsten. Qualitative Beurteilungskriterien sollten vermehrt bei PAs herangezogen werden, da hier eine translatorische Kraft am Wirbel appliziert wird und der Therapeut vom Gewebe mehr Feedback auf die eingebrachte Kraft erhalten kann.

2.4 Interrater-Reliabilität der passiven segmentalen Mobilitätstestung der Lendenwirbelsäule

Dieses Kapitel befasst sich anhand der jüngsten Reviews zu diesem Thema zuerst mit der Interrater-Reliabilität der passiven Mobilitätstestung im Allgemeinen. Danach wird auf die Qualitätsmängel vergangener Studien bzw. auf die Qualitätskriterien zukünftiger Studien eingegangen. Das letzte Unterkapitel behandelt nochmal die Interrater-Reliabilität der PPIVMs der LWS im Speziellen.

2.4.1 Reviews

In den letzten Jahren wurden einige narrative und systematische Reviews zur Reliabilität der Mobilitätstestung der Wirbelsäule veröffentlicht: Haneline et al. (2008); Haneline, Cooperstein, Young, und Birkeland (2009); Hestbaek und Leboeuf-Yde (2000); Huijbregts, (2002); May, Littlewood, und Bishop (2006); Seffinger et al., (2004); Stochkendahl et al. (2006); Triano et al. (2013); van Trijffel et al. (2005). Der Großteil kommt zusammenfassend zu dem Schluss, dass die passive segmentale Mobilitätstestung der Wirbelsäule eine schlechte Interrater-Reliabilität aufweist. Triano et al. und Seffinger et al. berichten auch von widersprüchlichen Ergebnissen zur Interrater-Reliabilität. Von Seffinger et al. und Haneline et al. (2008) wird die schwere Vergleichbarkeit von Studien aufgrund unterschiedlicher Palpationstechniken, unterschiedlicher statistischer Auswertungen und variierender Studiendesigns hervorgehoben. Haneline et al. (2008) unterteilten deshalb die Studien in ihrem Review nach Beurteilungskriterien der Palpation - mit dem Ergebnis, dass es keinen Einfluss auf die Interrater-Reliabilität hat, ob bei der passiven Untersuchung der Wirbelsäule nach qualitativen (Endgefühl) oder quantitativen (Bewegungsausmaß) Kriterien beurteilt wird. Nach Seffinger et al. wirft die schlechte Interrater-Reliabilität auch die Frage nach der Validität der Mobilitätstestung auf. Er gibt zu bedenken, dass, wenn bei der passiven segmentalen Mobilitätstestung der Wirbelsäule nicht das gemessen wird, was gemessen werden soll, eine gute Reliabilität dieser Testmethode sehr unwahrscheinlich sein wird.

Viele Autoren bemängeln aber vor allem die schlechte Qualität der bisherigen Interrater-Reliabilitätsstudien (Haneline et al., 2009; May et al., 2006; Seffinger et al., 2004; Stochkendahl et al., 2006; Triano et al., 2013; van Trijffel et al., 2005), speziell bei Studien, welche vor 1985 veröffentlicht wurden (Hestbaek & Leboeuf-Yde, 2000). In den letzten Jahren verbesserte sich die Studienqualität etwas (Seffinger et al., 2004).

Qualitätsbemängelungen betreffen folgende Punkte:

- Probanden
- Tester
- Studiendesign und Beschreibung
- Datenaufbereitung und statistische Methoden

Im folgenden Kapitel soll auf diese Punkte näher eingegangen werden und es sollen adäquate Ansätze zur Verbesserung der Methodik erhoben werden.

2.4.2 Qualitätskriterien

2.4.2.1 Probanden/Patienten

Asymptomatische, symptomatische oder gemischte Probandengruppe? Diese Frage stellt sich bei der Entwicklung des Studiendesigns sehr schnell. Huijbregts (2002), May et al. (2006) und Van Trijffel et al. (2005) bemängeln an vielen Studien, dass asymptotische Probanden untersucht wurden und diese nicht repräsentativ für die Praxis sind. Hingegen schreiben Seffinger et al. (2004) und Stochkendahl et al. (2006), dass die Durchführung der Studie an symptomatischen Patienten, d.h. repräsentativ für die Praxis, auch keinen Einfluss auf die Reliabilität hat.

Rückenschmerzen der Lendenwirbelsäule gehen oft mit somatischen Dysfunktionen im lumbalen, sacralen oder Becken-Bereich einher (Licciardone & Kearns, 2012). Durch die Manipulation der Lendenwirbelsäule können die Rückenschmerzen der LWS signifikant verringert werden (Licciardone, Brimhall, & King, 2005). Man könnte daraus schließen, dass asymptotische Probanden seltener oder keine Dysfunktionen der Lendenwirbel aufweisen und deshalb die Interrater-Reliabilität bei der Mobilitätstestung der LWS schlechter ausfällt. Um eine repräsentative Gruppe, wie sie in der täglichen Praxis vorkommt, zu erzielen, macht eine Mischung von symptomatischen (Rückenschmerzen) sowie asymptotischen (keine Rückenschmerzen, schließt jedoch nicht aus, dass andere Regionen von Schmerz betroffen sind) Probanden Sinn (Patijn & FIMM Scientific Committee, 2004).

May et al. (2006) bemängeln außerdem oft eine geringe Probandenzahl und die Art der Rekrutierung. Die Teilnehmer sollten randomisiert ausgewählt werden. Die Mindestteilnehmerzahl für Reliabilitätsstudien liegt laut FIMM - International Federation for Manual/Musculoskeletal Medicine (Patijn & FIMM Scientific Committee, 2004) bei 40 Probanden. Nach Bortz und Lienert (2003) ist für die statistische Auswertung nach Cohens Kappa eine Probandenzahl > 20 ausreichend. Die Reihenfolge, in welcher die Probanden von den Testern untersucht werden, sollte entweder alternierend erfolgen oder randomisiert werden (Lucas & Bogduk, 2011).

2.4.2.2 Tester

Blindierung: In Studien zur Interrater-Reliabilität sollen die Untersucher geblindet gegenüber den Probanden sowie den Ergebnissen der anderen Untersucher sein. Während der Untersuchung darf keine Kommunikation zwischen den Untersuchern bzw. zwischen Untersucher und Proband stattfinden. Die Aufzeichnung der Ergebnisse muss auf separaten Dokumenten erfolgen (Lucas & Bogduk, 2011; Patijn & FIMM Scientific Committee, 2004). Van Trijffel et al. (2005) kritisieren bei schlechten Studien die fehlende Blindung der Untersucher. May et al. (2006) schlagen zur Verbesserung der Qualität von Reliabilitätsstudien zusätzlich den Einsatz von unabhängigen Beobachtern vor. Ihre Aufgabe ist es zu gewährleisten, dass kein Informationsaustausch unter den Testern bzw. zwischen Testern und Probanden stattfindet.

Anzahl der Tester: Mehrere Tester-Paare sind laut May et al. (2006) für Reliabilitätsstudien notwendig. Dem widerspricht die FIMM (Patijn & FIMM Scientific Committee, 2004) mit dem Argument, dass es aus statistischer Sicht keine Gründe gibt, zur Ermittlung der Interrater-Reliabilität mehr als zwei Untersucher einzusetzen.

Therapeutenerfahrung: Bei bisherigen Interrater-Reliabilitätsstudien variiert naturgemäß die Erfahrung der Untersucher (siehe: Degenhardt et al., 2005; Deore & May, 2012; Maher & Adams, 1994; Qvistgaard et al., 2007). Eine Verbesserung der Interrater-Reliabilität durch erfahrenere Therapeuten konnte bisher jedoch nicht nachgewiesen werden (Seffinger et al., 2004; Stochkendahl et al., 2006). Die FIMM gibt sogar zu bedenken, dass erfahrenere Untersucher vermutlich persönliche Stile der Mobilitätstestung entwickelt haben und erachten es deshalb als wichtig, die Durchführung und Interpretation zu standardisieren (Patijn & FIMM Scientific Committee, 2004).

Training: Dass durch 4-monatiges Konsensus-Training die Interrater-Reliabilität verbessert werden kann, haben Degenhardt et al. (2010) belegt. Mit dem Training der Untersucher wird bei der statistischen Auswertung der Reliabilität aber zum Teil auch der Effekt des Trainings und nicht nur die Reliabilität der Technik selbst beurteilt (Huijbregts, 2002).

Zum einen sollte bei Reliabilitätsstudien die Mobilitätstestung der Lendenwirbelsäule so erfolgen, wie es der täglichen Praxis entspricht, zum anderen ist sie, wie oben erwähnt, durch persönliche Stile und verschiedene Interpretationsmöglichkeiten geprägt (s. Kapitel 2.3.2). Dies könnte auch ein Grund sein, warum die Intrarater-Reliabilität bei Mobilitätstestungen besser abschneidet als die Interrater-Reliabilität (Huijbregts, 2002). Um die Interrater-Reliabilität aussagekräftig vergleichen zu können, kommt man nicht umhin die Untersuchung zu standardisieren und die Therapeuten auf die Testung vorzubereiten. Eine Einigung der Untersucher über

das genaue Testprocedere und dessen Interpretation ist notwendig, da man sonst im schlimmsten Fall zwei unterschiedliche Tests vergleicht. Ein länger andauerndes Training entspricht jedoch nicht mehr der Praxisrealität und erscheint somit auch nicht sinnvoll.

Seffinger et al. (2004) und Stockendahl et al. (2006) fassen zusammen, dass eine längere Therapeutenerfahrung, Standardisierung der Testung, Training vor der Testung und Einbindung symptomatischer Probanden keine Auswirkung auf die Interrater-Reliabilität haben. Diese Aussagen sind jedoch kritisch zu betrachten, da die Qualität vieler Studien auch noch in vielen anderen Punkten bemängelt wird und dadurch möglicherweise falsche Schlüsse zu den genannten Komponenten gezogen wurden.

2.4.2.3 Studiendesign und Beschreibung

Studienprotokolle: In vielen Reviews zur Reliabilität der Mobilitätstestung der Wirbelsäule wird die mangelhafte Beschreibung der Studien kritisiert, sodass diese schwer reproduzierbar sind (Haneline et al., 2008; May et al., 2006; Seffinger et al., 2004; Stockendahl et al., 2006; van Trijffel et al., 2005). Dies betrifft v.a. die Beschreibung der Testung und der qualitativen oder quantitativen Beurteilungsparameter. Van Trijffel et al. schildern, dass in vielen Fällen keine Angaben zur Anzahl der Tests, Anzahl der Bewegungswiederholung, Kraftaufwendung in der Endposition, Bewegungsrichtung und zu den Zeitintervallen gemacht wurden.

Die Beurteilungsskalen, welche für Reliabilitätsstudien verwendet werden, sollen denen der täglichen Praxis entsprechen (Huijbregts, 2002). In der osteopathischen Untersuchung wird v.a. nach der Hypomobilität gesucht und eine dichotome Skala (hypomobil/nicht hypomobil) erscheint somit ausreichend.

Markierung der Dornfortsätze: Da bekannt ist, dass die Reliabilität bei der Benennung von Wirbelsegmenten in der LWS schlecht ist, kann sich das auch negativ auf die Interrater-Reliabilität bei der Mobilitätstestung auswirken (Triano et al., 2013). Die Tester befinden möglicherweise dasselbe Segment als hypomobil, aber identifizieren es als unterschiedliche Segmente (Haneline et al., 2008). Die Markierung der Dornfortsätze könnte diesem Fehler entgegenwirken, was bei jüngeren Studien auch vermehrt gemacht wurde, damit die Tester von denselben Wirbelsegmenten sprechen (siehe: Degenhardt et al., 2010; Deore & May, 2012; Qvistgaard et al., 2007; Schneider et al., 2008). Haneline et al. (2008) geben dabei zu bedenken, dass dies die Interrater-Reliabilität künstlich verbessern würde, da in der Praxis keine Markierungen zur Anwendung kommen.

Bisherige Studien, bei denen die Wirbelsegmente markiert wurden, zeigten auch keine Verbesserung der Interrater-Reliabilität (van Trijffel et al., 2005). Aber auch hier erscheint es der Autorin wichtig zu erwähnen, dass eine Kombination methodologischer Unzulänglichkeiten bei früheren Studien zu fehlerhaften Rückschlüssen führen könnte. Um die Vergleichbarkeit mit anderen Studien zu erleichtern und sicherzustellen, dass die Tester von denselben Segmenten sprechen, ist die Markierung der Dornfortsätze unerlässlich.

Mobilisierende Effekte: Obwohl die biomechanischen Eigenschaften des Bindegewebes eine zentrale Rolle bei der Mobilitätsuntersuchung spielen, wird dieses Thema in keiner Studie explizit behandelt (van Trijffel et al., 2005). Diese sind aber anfällig für Veränderungen aufgrund mobilisierender Effekte und könnten somit ebenfalls eine Fehlerquelle in Interrater-Reliabilitätsstudien darstellen.

In einer biomechanischen Studie von Bisschop et al. (2013) zum Effekt des wiederholten Bewegens auf die Mobilität der Wirbelsäule wird eine signifikante Verbesserung des Bewegungsausmaßes angegeben. Die Studie wurde in vitro an Präparaten der Wirbelsäule durchgeführt, die Bewegungen wurden in jede Bewegungsrichtung zehnmal wiederholt. Obwohl sich die Beweglichkeit in alle Bewegungsrichtungen verbessert hat, liegt das Ausmaß in allen Ebenen unter 1°. Dieses Erkenntnis ist auf lebende Probanden nur bedingt übertragbar, da viele Faktoren wie z.B. Muskelspannung oder neurophysiologische Komponenten einen Einfluss auf die Mobilität haben. Zum anderen darf in Frage gestellt werden, ob die Tester in der Lage sind, Veränderungen <1° zu palpieren. Demzufolge scheinen die mobilisierenden Effekte bei zehn Wiederholungen vernachlässigbar.

2.4.2.4 Datenaufbereitung und statistische Methoden

Datenaufbereitung: Qualitätsmängel bei der Datenanalyse oder Präsentation der Ergebnisse wurden von Seffinger et al. (2004) und Stockendahl et al. (2006) kritisiert. Van Trijffel et al. (2005) empfehlen bei der Verwendung von Kappa-Statistiken die Datenaufbereitung für den Leser in Kreuztabellen.

Statistische Methoden: Für Interrater-Reliabilitätsstudien wird die Verwendung von Cohens Kappa zur statistischen Auswertung empfohlen (Lucas & Bogduk, 2011; Patijn & FIMM Scientific Committee, 2004). In früheren Studien kamen immer wieder inadäquate statistische Methoden zum Einsatz, wie z.B. Angaben in prozentueller Übereinkunft (Huijbregts, 2002; May et al., 2006; Stockendahl et al., 2006; van Trijffel et al., 2005). Das Problem dabei ist, dass Übereinkünfte, welche durch Zufall zu erwarten sind, statistisch nicht beachtet werden (Patijn & FIMM Scientific Committee, 2004).

2.4.2.5 Zusammenfassung der Qualitätskriterien für Interrater-Reliabilitätsstudien

In Kapitel 2.4.2 beschriebene Qualitätskriterien zur Verbesserung von Reliabilitätsstudien sollen hier noch einmal übersichtlich dargestellt werden:

- Klinisch repräsentative Probanden (d.h. symptomatische UND asymptomatische Probanden)
- Patientenzahl > 20
- Probanden randomisiert rekrutiert
- Durchführung der Testung alternierend oder randomisiert
- Blindung der Untersucher gegenüber Probanden und Ergebnissen der anderen Untersucher
- Kontrolle durch unabhängige Beobachter
- Standardisierung von Testprozedere und Interpretation der Ergebnisse
- Ausführliche Studienprotokolle
- Markierung der Dornfortsätze
- Statistische Auswertung mittels Cohens Kappa

2.4.3 Interrater-Reliabilität der spezifisch segmentalen physiologischen Bewegungen (PPIVMs) der Lendenwirbelsäule

Wie in Kapitel 2.4.1 bereits zusammengefasst, ist die Interrater-Reliabilität der segmentalen passiven Mobilitätstestung der Wirbelsäule schlecht. Dieses Erkenntnis bezieht sich allerdings auf die gesamte Wirbelsäule, ohne zwischen PAs oder PPIVMs zu unterscheiden. Insgesamt finden sich in der jüngeren Literatur deutlich mehr Studien, welche die PAs untersuchten (z.B.: Binkley et al., 1995; Degenhardt et al., 2005; Deore & May, 2012; Horneij, Hemborg, Johnsson, & Ekdahl, 2002; Maher & Adams, 1994; Schneider et al., 2008; Shaw et al., 2012) als Studien zur Untersuchung der PPIVMs.

Frühere Studien, welche die PPIVMs der LWS im Sitz testeten, wiesen immer wieder qualitative Mängel auf, indem sie entweder nur asymptomatische Probanden untersuchten, keine Markierungen der Dornfortsätze aufwiesen oder ungeeignete statistische Auswertungsmethoden anwendeten (Haneline et al., 2009; Huijbregts, 2002). In einer adaptierten Tabelle von Haneline et al. (2009) werden zur besseren Vergleichbarkeit die Interrater-Reliabilitätsstudien, welche an der LWS im Sitz durchgeführt wurden, aufgelistet:

Tabelle 1. Interrater-Reliabilitätsstudien der LWS im Sitz; Quelle: (adaptiert) Haneline et al.,(2009)

Author, Bibliography	Region	Examiners, experience	Subjects	Quality Score	Findings	Degree of Reliability
Bergstrom & Courtis (1986)	L1-L5	2 DC	100 Asx	67%	% = 65 - 88	Inconclusive
Keating et al. (1990)	T12-S1	3 DC, >2,5 yrs	46 (21Sx, 25 Asx)	67%	$k = -0,18$ to $0,31$	None to fair
Love & Brodeur (1987)	T1-L5	8 DC St	32 Asx	17%	$r = 0,01$ to $0,49$	Inconclusive
Mootz et al. (1989)	L1-S1	2 DC, ≥ 7	60 Asx	33%	$k = -0,17$ to $0,17$	None to slight

DC = doctor of chiropractic; St = students; Asx = asymptomatic; Sx = symptomatic; r = Pearsons correlation coefficient

In der gesamten von Haneline et al. gesammelten Bibliographie finden sich nur vier Studien, welche die PPIVMs im Sitz testeten. Alle vier wurden bereits zwischen 1986 und 1990 durchgeführt und sind somit schon älter. Zur Beurteilung der methodologischen Qualität wurde von Haneline et al. eine 6-Punkte-Skala nach Stochkendahl et al. (2006) herangezogen. Ob ihrer mangelhaft methodologischen Qualität sind diese oft kaum vergleichbar. Nur Keating et al. (1990) und Mootz et al. (1989) verwendeten zur statistischen Auswertung Cohens Kappa und erreichten Werte zwischen $k = -0,18$ und $0,31$ - das heißt, dass die Ergebnisse zum Teil unter

der Zufallsübereinstimmung lagen. Der Quality Score von Mootz et al. ist mit 33% jedoch relativ niedrig und bei Keating et al. fehlte die Markierung der Dornfortsätze. Bergstrom und Courtis (1986) errechneten die Interrater-Reliabilität prozentuell und beschrieben eine Übereinstimmung von 65 - 88%. Love und Brodeur (1987) errechneten die Interrater-Reliabilität mit Pearsons correlation coefficient mit $r = 0,01 - 0,49$.

In einem Review zur qualitativen vs. quantitativen Mobilitätstestung der Wirbelsäule von Haneline et al. (2008) werden folgende Studien, welche die Mobilität der LWS quantitativ beurteilten, d.h. das Bewegungsausmaß, zitiert (Tab. 2): Strender, Sjöblom, Sundell, Ludwig, & Taube (1997) erreichen bei der Interrater-Reliabilitätstestung der PPIVMs in der LWS in Seitenlage Kappa-Werte zwischen -0,08 und 0,75. Der Quality Score beträgt 67%. Lundberg und Gerdle (1999) untersuchten die PPIVMs ebenfalls in Seitenlage von L4/L5 und L5/S1 und errechneten den Weighted-Kappa mit $k_w = 0,59$ bis 0,75. Der Quality Score lag bei dieser Studie bei nur 50%. Mit einem niedrigen Quality Score von nur 17% erreichten Inscoe, Witt, Gross, & Mitchell (1995) bei der Untersuchung von Th12 - S1 eine prozentuelle Übereinstimmung von % = 33,3 - 58,3 (Scott's Pi = 18,4%). Phillips und Twomey (1996) erzielten bei einem Quality-Score von 67% Weighted-Kappa Werte von $k_w = -0,11$ bis 0,32.

Tabelle 2. Interrater-Reliabilitätsstudien der LWS beurteilt nach Bewegungsausmaß; Quelle: (adaptiert) Haneline et al. (2008)

Author, Bibliography	Region	Examiners, experience	Subjects	Quality Score	Findings	Degree of Reliability
Inscoe et al.(1995)	T12-S1	2 PT, $\geq 4y$	6 Sx	17%	Scott's Pi =18,4% % =33,3 - 58,3	Not acceptable
Lundberg & Gerdle (1999)	Th10-S1	3 PT, exp	150 Asx	50%	$k_w = 0,59$ to 0,75	Moderate to substantial
Phillips & Twomey (1996)	L1-5	2 PT, NI	72 (63 Sx, 9 Asx)	67%	$k_w = -0,11$ to 0,32	None to fair
Strender et al. (1997)	L5-S1	2 MD, 2 PT, exp	71 Sx	67%	$k = -0,08$ to 0,75 % = 48 - 88	None to substantial

PT = physical therapist; y = year; exp = experienced; MD = doctor of medicine; Sx = symptomatic; Asx = asymptomatic; % percent agreement; k = kappa; k_w = weighted kappa; NI = (k. a.)

Diese Studien wurden alle vor dem Jahr 2000 durchgeführt. Danach wurden nur mehr wenige Studien, welche die Interrater-Reliabilität der PPIVMs an der Lendenwirbelsäule testeten, veröffentlicht:

Johansson veröffentlichte 2006 eine Interrater-Reliabilitätsstudie der LWS in Seitenlage. Drei Physiotherapeuten untersuchten die Flexion und Extension von L1 bis L5 an 20 sowohl symptomatischen als auch asymptomatischen Probanden. Die Kappa-Werte wurden pro Segment

berechnet. In 24 aus 30 Vergleichen blieb der Kappa-Wert unter 0,20. In vier Vergleichen lag er zwischen 0,21 und 0,40 und nur zweimal lag er darüber. Die Kappa-Werte für die Flexion reichten von $k = -0,18$ bis 0,21 ($k_w = -0,14$ bis 0,28) und von $k = -0,25$ bis 0,53 ($k_w = -0,09$ bis 0,56) für die Extension.

Qvistgaard et al. (2007) wiesen bereits ein besseres methodisches Studiendesign auf (Symptomatische Probanden, Markierung der Segmente, unabhängige Beobachter, geblindete Untersucher, statistische Auswertung nach Cohens Kappa). In der 2007 durchgeführten Studie kamen verschiedene Tests zur Anwendung, mit dem Ziel am Ende ein Segment mit Dysfunktion zu benennen. Dazu wurden 60 sowohl symptomatische als auch asymptomatische Probanden untersucht. Bei der Untersuchung der PPIVMs in Seitenlage ergab die statistische Auswertung für die Extension $k = 0,23$. Nach der Interpretation von Landis und Koch (1977; zitiert nach Sim & Wright, 2005) entspricht das einer „ausreichenden“ Übereinstimmung (siehe Seite 37). Qvistgaard et al. errechneten zusätzlich die Kappa-Werte für eine sogenannte „akzeptable Übereinstimmung der Untersucher“ - d.h., wenn die Untersucher zwei benachbarte Segmente als hypomobil gefunden hatten, wurde das als positives Ergebnis gewertet. In diesem Fall führte die Berechnung des Kappa-Koeffizienten zu einem verbesserten Ergebnis von $k = 0,52$, was nach Landis und Koch bereits als „mittelmäßige“ Übereinstimmung betrachtet werden kann.

Insgesamt konnte festgestellt werden, dass es nur wenige qualitativ gute Studien zur Untersuchung der PPIVMs in der Lendenwirbelsäule in Seitenlage gibt, welche maximal eine „ausreichende“ Reliabilität erreichten. Für die Untersuchung im Sitz fehlen diese überhaupt und sämtliche Ergebnisse beruhen auf mangelhaften Studien. Wie bereits erwähnt, sind diese Werte nur bedingt vergleichbar aufgrund der oft unterschiedlich methodischen Qualität. Weiter wird auch ersichtlich, dass mit zunehmender Qualität der Studien die schlechte Interrater-Reliabilität auch keine Verbesserung erreicht.

3 FORSCHUNGSFRAGE und HYPOTHESEN

Das Ziel dieser Studie ist es herauszufinden, ob die Untersuchung der Mobilität der einzelnen Segmente der Lendenwirbelsäule im Sitz sowohl an symptomatischen als auch an asymptomatischen Patienten verlässliche, nachvollziehbare Ergebnisse bringt. Weiters soll erforscht werden, ob die Interrater-Reliabilität bei den unterschiedlichen Bewegungskomponenten variiert.

Dazu werden folgende Hypothesen formuliert:

NULLHYPOTHESE 1:

Die palpatorischen Ergebnisse der segmentalen passiven Mobilitätstestung in den drei Bewegungsebenen der Lendenwirbelsäule im Sitz sind abhängig vom Untersucher.

Nullhypothese 1a: in der Sagittalebene

Nullhypothese 1b: in der Frontalebene

Nullhypothese 1c: in der Transversalebene

ALTERNATIVHYPOTHESE 1

Die palpatorischen Ergebnisse bei der segmentalen passiven Mobilitätstestung in den drei Bewegungsebenen der Lendenwirbelsäule sind reproduzierbar und unabhängig vom Untersucher.

Alternativhypothese 1a: in der Sagittalebene

Alternativhypothese 1b: in der Frontalebene

Alternativhypothese 1c: in der Transversalebene

NULLHYPOTHESE 2

Die palpatorischen Ergebnisse der segmentalen passiven Mobilitätstestung in den verschiedenen Bewegungsrichtungen der Lendenwirbelsäule im Sitz sind abhängig vom Untersucher.

Nullhypothese 2a: in Flexion

Nullhypothese 2b: in Extension

Nullhypothese 2c: in Lateralflexion links

Nullhypothese 2d: in Lateralflexion rechts

Nullhypothese 2e: in Rotation links

Nullhypothese 2f: in Rotation rechts

ALTERNATIVHYPOTHESE 2

Die palpatorischen Ergebnisse bei der segmentalen passiven Mobilitätstestung in den verschiedenen Bewegungsrichtungen der Lendenwirbelsäule sind reproduzierbar und unabhängig vom Untersucher.

Alternativhypothese 2a: in Flexion

Alternativhypothese 2b: in Extension

Alternativhypothese 2c: in Lateralflexion links

Alternativhypothese 2d: in Lateralflexion rechts

Alternativhypothese 2e: in Rotation links

Alternativhypothese 2f: in Rotation rechts

4 METHODIK

4.1 Literaturrecherche und Datenbanken

Die Erhebung der aktuellen Datenlage sowie das Auffinden relevanter Literatur zu den Themen Reliabilitätsstudien, Mobilitätstestung und Statistik in der klinischen Forschung erfolgte zum einen über Online-Recherche in diversen medizinischen Datenbanken und osteopathischen, physiotherapeutischen oder manualtherapeutischen Journalen - zum anderen wurden anhand der Literaturverzeichnisse der gefundenen Studien weitere relevante Artikel und Bücher aufgesucht.

Die Online-Literatursuche fand primär innerhalb von vier Tagen zwischen 5. und 8. März 2014 statt. Im Laufe des Durchlesens der Studien und bei der Durchsicht der Literaturangaben kam auch später noch relevante Literatur dazu.

In erster Linie erfolgte die Suche über das Portal PubMed (<http://www.ncbi.nlm.nih.gov/pubmed>), in welchem auch die meisten geeigneten Ergebnisse erzielt wurden. Zusätzlich wurde in nachstehenden Portalen und Journalen recherchiert, wobei für diese Studie hier nur mehr vereinzelt bedeutsame Artikel aufzufinden waren:

- The Cochrane library: www.thecochranelibrary.com
- International Journal of Osteopathic Medicine: www.journalofosteopathicmedicine.com
- Osteopathic research web: www.osteopathic-research.com
- The Journal of the American Osteopathic Association (JAOA): www.jaoa.org
- OSTMED.DR - osteopathic medicine digital library: www.ostmed-dr.com
- ClinicalKey - Elsevier: www.clinicalkey.com
- Chiropractic & Manual Therapies: www.chiromt.com
- PEDro - Physiotherapie Evidenz Datenbank: www.pedro.org.au
- Journal of Bodywork and Movement Therapies: www.bodyworkmovementtherapies.com
- Journal of the American Physical Therapy Association: <http://ptjournal.apta.org>
- Journal of Manipulative and Physiological Therapeutics: www.jmptonline.org
- Journal of Physiotherapy: www.journalofphysiotherapy.com

In unterschiedlichsten Kombinationen wurden folgende Keywords verwendet: lumbar, spine/spinal, motion, mobility, segmental, manual examination, interrater/interexaminer, reliability, palpation/palpatory, diagnostic.

4.2 Studiendesign

Bei der vorliegenden Studie handelt es sich um eine methodologische Studie.

Zur Erhebung der Interrater-Reliabilität der passiven segmentalen Mobilitätstestung der Lendenwirbelsäule wurden sowohl symptomatische als auch asymptotische Probanden rekrutiert. Nach Kennzeichnung der Processus spinosi der LWS erfolgte die Untersuchung durch zwei erfahrene Therapeuten. Die Ergebnisse wurden verglichen, statistisch ausgewertet und interpretiert.

4.3 Die Untersucher

Für die Durchführung der Untersuchung stellten sich zwei Kollegen der Studienleiterin zur Verfügung. Beide Tester befanden sich gemeinsam mit der Studienleiterin in Ausbildung an der Wiener Schule für Osteopathie im 6. Ausbildungsjahr/3. Semester Universitäts-Lehrgang „Osteopathie MSc“ in Kooperation mit der Donau-Uni Krems.

Testerin 1 war 40 Jahre alt und hatte 12 Jahre Berufserfahrung als Physiotherapeutin. Sie wendete die segmentale Untersuchung der Wirbelsäule seit Beginn der Osteopathie-Ausbildung 2008 regelmäßig an.

Tester 2 war 42 Jahre alt und hatte 13 Jahre Berufserfahrung als Physiotherapeut. Auch er wendete die segmentale Untersuchung der Wirbelsäule seit Beginn der Osteopathie-Ausbildung 2008 regelmäßig an.

4.4 Die Assistentinnen

Beide Untersucher wurden jeweils von einer Assistentin begleitet. Die Assistentinnen waren eine Physiotherapeutin mit 4 Jahren Berufserfahrung und eine Physiotherapeutin/Osteopathin i.A. mit 8 Jahren Berufserfahrung. Ihre Aufgabe bestand einerseits darin die Blindierung zu gewährleisten, indem sie die Kommunikation zwischen Tester/Proband und Tester/Tester verfolgten. Andererseits übernahmen sie die Dokumentation der Testergebnisse und unterstützen damit einen flüssigen Ablauf der Untersuchung, sodass die Tester ohne Unterbrechung arbeiten konnten.

4.5 Die Probanden

Um eine möglichst praxisnahe Situation herzustellen, wurden sowohl Probanden mit als auch ohne Rückenschmerzen rekrutiert. Die Rekrutierung konzentrierte sich aber v.a. auf Patienten mit Rückenschmerzen in der Lendenwirbelsäule. Das Mindestalter betrug 18 Jahre, damit die Entwicklung der Wirbelsäule bei allen weitgehend abgeschlossen war und eine ähnliche Ausgangssituation für alle Probanden herrschte.

Die Ausschlusskriterien waren wie folgt:

- Operationen im Rumpf-/Hüft-/Armbereich in den letzten 6 Wochen
- Patienten, die aufgrund akuter Schmerzen nicht sitzen können
- Akuter Diskusprolaps mit radikulärer Symptomatik
- Schwangerschaft

Operationen im Rumpf-/Hüft- oder Armbereich in den letzten 6 Wochen könnten die benötigte Bewegungsfreiheit limitieren bzw. könnten durch das wiederholte Bewegen erneute Reizzustände hervorgerufen werden. Patienten, die aufgrund akuter Schmerzen nicht sitzen können, wurden aus der Studie ausgeschlossen, da der Sitz die Ausgangsposition der Untersuchung darstellte. Für Patienten mit akuter radikulärer Symptomatik könnte die Untersuchung eine erneute Nervenreizung provozieren und die bestehende Symptomatik verstärken. Mobilisierende Techniken an der LWS sind im ersten Trimenon der Schwangerschaft aufgrund der erhöhten Gefahr eines Abortes, kontraindiziert und mit fortschreitender Schwangerschaft kommt es zur Auflockerung des Bindegewebes. Da nicht gewährleistet werden kann, dass die Palpation einen mobilisierenden Effekt hat, galt Schwangerschaft ebenfalls als Ausschlusskriterium.

Um für die statistische Auswertung genügend Probanden (>20) zu erreichen, wurde die Teilnahme von 24 Personen angestrebt. Die Rekrutierung erfolgte via E-Mail und SMS im Freundes-/Verwandten- und Bekanntenkreis der Studienleiterin mit der Bitte um Weiterleitung. 26 Personen meldeten sich freiwillig. Einer wurde aufgrund einer radikulären Symptomatik ausgeschlossen und eine wegen Schwangerschaft. Die Probanden wurden vorab via E-Mail mit einem Informationsblatt zu Inhalt und Ziel der Studie informiert und unterzeichneten eine Einverständniserklärung (siehe Anhang). Sie wurden außerdem darüber aufgeklärt, dass dies nur eine Untersuchung darstelle und keine Behandlung beinhalte. Als Entschädigung für den entstandenen Zeitaufwand erhielt jeder Teilnehmer einen Massagegutschein für die Dauer von 45 min.

Durch einen Ausfall wegen Krankheit und einem, weil der Termin vergessen wurde, nahmen schlussendlich 22 Probanden (n=22), davon 12 Frauen und 10 Männer, teil. Das Alter der teilnehmenden Probanden betrug durchschnittlich $32,6 \pm 5,9$ Jahre (Min. = 24, Max. = 51, Median = 31). Die demographischen Daten der Probanden werden in Tabelle 3 dargestellt.

Tabelle 3. Demographische Verteilung der Probanden

Patientencharakteristika (n=22, w=12, m=10)			
	Alter (Jahre)	Größe (cm)	Gewicht (kg)
Arithmet. Mittel	32,6	174,8	75,2
SD	5,9	8,7	13,0
Median	31	175	75
Min.	24	161	56
Max.	51	189	100

Acht Probanden hatten zum Zeitpunkt der Untersuchung Schmerzen in der LWS. Vierzehn Probanden waren zum Zeitpunkt der Untersuchung zwar schmerzfrei, davon gaben aber neun innerhalb der letzten sechs Wochen Schmerzen in der Lendenwirbelsäule an, wie aus Tabelle 4 ersichtlich wird. Fünf Probanden waren in der Lendenwirbelsäule asymptomatisch, was aber Schmerzen in Hals- bzw. Brustwirbelsäule oder Nacken nicht ausschließt.

Tabelle 4. Schmerzsituation LWS der Probanden

Schmerzsituation LWS	n	%
Aktuell Schmerzen	8	36%
Innerhalb der letzten 6 Wochen	9	41%
Schmerzfrei	5	23%
gesamt	22	100%

4.6 Blindierung

Proband - Tester - Verblindung: Die Probanden waren den Testern völlig unbekannt. Die Tester waren geblindet bezüglich der Klinik bzw. Symptomatik der Probanden. Um einen Informationsaustausch über die Symptomatik zu vermeiden, wurden die Tester und Probanden angewiesen nicht zu reden. Ausnahmen bestanden für die Begrüßung oder notwendige Anweisungen für die Untersuchung (Änderung der Sitzposition, Haltung, locker lassen). Um dies zu gewährleisten, wurde jeder Tester von einer Assistenzperson begleitet.

Tester - Tester - Verblindung: Da die Tester nacheinander die Untersuchungsräume betraten und sich immer wieder am Gang oder im Aufenthaltsraum trafen, waren sie angewiesen, keine Information über die Probanden und die Ergebnisse der Untersuchung auszutauschen. Dies wurde ebenfalls durch die Assistenzpersonen gewährleistet.

4.7 Die Untersuchung

4.7.1 Tester-Abstimmung

Dreieinhalb Wochen vor der eigentlichen Durchführung der Studie wurden der gesamte Ablauf mit den Testern und einer Assistentin besprochen und die Testung selbst abgestimmt. Es wurden die Handhaltung der palpierenden Hand diskutiert und festgelegt, die Beurteilungskriterien erläutert und festgelegt und die richtige Dokumentation beschrieben. Beide Tester mussten sich in diesen Punkten einig sein, um eine standardisierte Mobilitätstestung zu gewährleisten. Zusätzlich wurde die Testung mehrmals durchgemacht und dabei die Zeit genommen, um einen ungefähren Zeitplan erstellen zu können.

4.7.2 Durchführungsort

Die Durchführung der Studie fand in der Praxis der Studienleiterin statt, einer privaten Gemeinschaftspraxis in Linz (OÖ). Dort standen für die Untersuchung drei abgetrennte Räume zur Verfügung, welche die Tester nacheinander zirkulierend betraten.

Der ursprüngliche Plan, dass die Tester immer im selben Raum bleiben und die Probanden von Raum zu Raum wechseln sollten, wurde verworfen. Das Aufstehen, An- und Auskleiden und Gehen der Probanden zwischen den Untersuchungen hätte die Mobilität der Wirbelsäule oder Kennzeichnungen an den Dornfortsätzen beeinflussen können. Da die Blindierung der Tester ohnehin durch die Assistenten kontrolliert wurde, konnte es in Kauf genommen werden, dass sich die Tester zwischen den Untersuchungen treffen.

4.7.3 Ablauf

Die Studie wurde am 3. Mai 2014 durchgeführt. Die Einteilung der Probanden erfolgte willkürlich. Dabei wurden auch Terminwünsche berücksichtigt. Pro 30 Minuten wurden drei Probanden eingeteilt und nach Reihenfolge des Eintreffens den drei Räumen zugeteilt. Die Studienleiterin selbst betrat den Raum zuerst. Es erfolgten das Ausfüllen des Fragebogens und die Unterzeichnung der Einverständniserklärung. Bei Bedarf wurden noch offene Fragen geklärt und die Patienten erhielten eine Nummer in Form eines Aufklebers am Oberarm. Diese Nummer wurde später am Dokumentationsblatt eingetragen, um die Ergebnisse den Teilnehmern zuzuordnen zu können. Danach markierte die Studienleiterin die Dornfortsätze der Lendenwirbelsäule und auf Wunsch der Tester auch den ersten Sakralwirbel (siehe Kapitel 4.7.5).

Darauffolgend wurde die Mobilitätstestung durchgeführt. Erst nachdem der erste Tester fertig war, durfte der zweite Tester eintreten. Pro Proband hatten die Tester ca. 10 Minuten Zeit, um die gesamte Mobilitätstestung durchzuführen. Diese Zeit stellte sich als ausreichend heraus, sodass keine Hektik entstand. Nach 12 Probanden wurde eine 30-minütige Pause eingeplant.

Testreihenfolge: Vor der Pause erfolgte die Mobilitätstestung immer von Tester 2 zuerst (n=12; 55%). Nach der Pause wurde die Reihenfolge umgedreht und Tester 1 untersuchte immer als Erster (n=10; 45%).

4.7.4 Fragebogen und Materialien

Die Untersuchung erfolgte rein manuell, somit waren keine anderen Messinstrumente für die Mobilitätstestung im Einsatz.

Vor der Untersuchung wurde ein Fragebogen ausgefüllt. Dieser diente der Erhebung der symptomatischen und asymptomatischen Teilnehmer. Die Studienleiterin ging mit den Probanden gemeinsam den Fragebogen durch und füllte ihn persönlich aus. Der Fragebogen selbst war kurz gehalten und beinhaltete als Erstes die Frage nach dem momentanen Schmerzzustand der Lendenwirbelsäule. Auf die Beantwortung der Folgefrage, wie lange die Schmerzen schon bestünden, falls vorhanden, hatte die Studienleiterin in Ihrer Nervosität in den meisten Fällen vergessen und somit wird dieser Punkt nicht bewertet.

Die zweite Frage richtete sich an die Probanden, die aktuell keine Schmerzen in der Lendenwirbelsäule hatten und diente der Erhebung dessen, ob es in den letzten sechs Wochen Schmerzsituationen gab. Damit fielen die Probanden, welche mit „Ja“ antworteten, ebenfalls in die Kategorie der symptomatischen Probanden.

Insgesamt hatten 8 Probanden am Untersuchungstag aktuell Schmerzen in der Lendenwirbelsäule. Von den 14 Probanden, welche zum Untersuchungszeitpunkt schmerzfrei waren, gaben 9 Probanden innerhalb der letzten sechs Wochen Schmerzen im Lumbalbereich an (siehe Tabelle 4, Seite 28).

Für die Kennzeichnung der Dornfortsätze war die Verwendung eines wasserlöslichen Stiftes, der im Trockenzustand nicht verwischt, geplant, welcher vorab an der Hand getestet wurde. Da aber schon bei der ersten Probandin aufgrund der starken Hautfeuchtigkeit im lumbalen Bereich die Markierungen verwischten, wurde kurzerhand ein Kugelschreiber verwendet. Dieser hatte den Nachteil, dass nach wiederholtem Palpieren die Markierungen schlechter sichtbar wurden.

4.7.5 Markierung der Processus spinosi der LWS

Aufgrund der schlechten Reliabilität bei der Benennung der lumbalen Segmente wurden vor der Untersuchung die Processus spinosi der Lendenwirbelsäule - und auf Wunsch der Tester auch des ersten Sakralwirbels - von der Studienleiterin markiert (Abb. 1).



Abbildung 1. Markierung der Processus spinosi

Der 3. Lendenwirbel wurde beschriftet. Snider et al. (2011) stellten fest, dass sich die palpatorische Treffsicherheit für L1 - L4 verbessert, wenn man sich dazu an folgenden knöchernen Fixpunkten orientiert:

- Identifikation von Th12 durch den kleineren Processus spinosus im Vergleich zu L1, um die Höhe von L1 aufzufinden;
- Identifikation der 12. Rippen und ihren Ansätzen an Th12, um Th12 zu identifizieren und folglich den Proc. spinosus von L1;
- Identifikation der Beckenkämme und Auffinden von L4 auf deren Verbindungslinie;
- Identifikation der Sakralbasis und folglich des Proc. spinosus L5;
- Identifikation des Proc. spinosus von L5 durch die kleinere Größe des Dornfortsatzes und folglich auch des Proc. spinosus L4;

Die Markierung wurde an den sitzenden Probanden eingezeichnet. Die Lendenwirbelsäule befand sich möglichst in einer Neutralposition. Das Einzeichnen gestaltete sich schwierig, da im Sitz auch die Spannung des umliegenden Gewebes erhöht ist. Teilweise waren die Zwischenräume der Dornfortsätze nicht palpierbar. Deshalb wurde bei Bedarf zusätzlich eine kleine Bewegung über den Rumpf induziert, um durch das Öffnen und Schließen der Procc. spinosi die Zwischenräume palpieren zu können. Eine besondere Schwierigkeit stellte die Palpation des Proc. spinosus von L5 dar, v.a. auch bei der Unterscheidung von L5/S1/S2.

4.7.6 Durchführung der passiven segmentalen physiologischen Mobilitätstestung der Lendenwirbelsäule

Für die osteopathische Befunderhebung der Wirbelsäule wird gewöhnlich nach den hypomobilen Segmenten gesucht. Üblicherweise erfolgt die Interpretation der gefundenen Ergebnisse erst nach einer Reihe von Untersuchungen. Zu Beginn wird eine globale Untersuchung ausgeführt, welche dann in den „auffälligen“ Regionen eine segmentale Untersuchung nach sich zieht. Diese wiederum „scannt“ die Wirbel in eine Bewegungsrichtung durch und bestimmt ebenfalls das „auffälligste“ bzw. „hypomobilste“ Segment. Ist das gefunden, wird dieses Segment auch noch in den anderen Bewegungsrichtungen untersucht und diese Ergebnisse dann interpretiert.

In dieser Studie waren die Tester angewiesen, die verschiedenen Bewegungsrichtungen segmental zu untersuchen und das „hypomobilste“ Segment pro Bewegungsrichtung zu bestimmen. Eine weitere Interpretation der Ergebnisse oder Überlegungen, wie das Gefundene zusammenpasst, wurde ausdrücklich nicht gefordert. Es ging rein darum herauszufinden, ob eine Interrater-Reliabilität bei der segmentalen physiologischen Mobilitätstestung besteht.

Dazu wurden die Probanden wie folgt untersucht:

AUSGANGSSTELLUNG: Für die Testung wurde als Ausgangsposition der Sitz gewählt. Einerseits um Positionswechsel für die Untersuchung der Lateralflexion und Rotation zu vermeiden; andererseits weil die Untersuchung im Sitz eine gängige Methode ist und in der Praxis oft Anwendung findet. Die Arme wurden vor dem Körper verschränkt - mit den Händen auf den Schultern. Die Probanden saßen am Ende der Liege, sodass die Therapeuten bei Bedarf die Seiten wechseln konnten.

POSITION DER TESTER: Die Therapeuten standen seitlich der Probanden. Ein Arm umgriff den Oberkörper von ventral und führte den Probanden damit (Abb.2). Die andere Hand lag dorsal und palpierter mit den Fingerkuppen zwischen den Procc. spinosi der LWS (Abb.1). Auf welcher Seite die Tester standen und mit welcher Hand sie palpieren, war den Testern überlassen. Ein Seitenwechsel war auch während der Testung jederzeit möglich.



Abbildung 2. Testerposition und Handhaltung

TEST und BEURTEILUNG: Während mit der ventralen Hand über den Oberkörper eine Flexion, Extension, Lateralflexion oder Rotation in der Lendenwirbelsäule induziert wurde, palpierter die dorsale Hand mit den Fingerkuppen zwischen den Procc. spinosi. Es wurde auf das Öffnen und Schließen zwischen den Wirbeln geachtet. Die Segmente wurden nach quantitativen Kriterien beurteilt (siehe Kapitel 2.3.2). Die Therapeuten wurden mit folgenden praxisüblichen Fragen in die Testung geschickt: *Welches der untersuchten Segmente ist das hypomobilste? bzw. Welches Segment bedarf deshalb einer genaueren Untersuchung?* Während der Palpation

aller fünf Segmente war eine Entscheidung anhand einer dichotomen Skala mit den Möglichkeiten „hypomobil/nicht hypomobil“ zu treffen. Die intersegmentale Bewegung sollte nicht öfter als 2- bis 3-mal durchgeführt werden, um einen möglichen Mobilisationseffekt so gering wie möglich zu halten. Am Ende jeder Untersuchungsrichtung war das hypomobilste Segment zu bestimmen. Die Assistentinnen kontrollierten den Ablauf und trugen die Ergebnisse in dafür vorgefertigte Befundtabellen ein, welche anschließend in einem Kuvert gesammelt wurden (siehe Anhang).

4.7.7 Feedback der Tester direkt nach der Testung

Nach der Testung wurde ein Feedback der Tester eingeholt, um eventuelle methodische Fehler oder andere Fehler in Folgestudien vermeiden zu können. Die Schwierigkeiten betrafen vor allem zwei Komponenten: die Palpation und die Interpretation.

PALPATION:

Die Palpation zwischen den Dornfortsätzen war nicht immer wie geplant durchführbar, da dadurch teilweise zu wenig Information gewonnen werden konnte. Ein Tester gab an, dass aufgrund der höheren Gewebespannung v.a. die Lendenwirbelsäule der männlichen Probanden oft schwerer zu palpieren war. Deshalb wendeten die Tester, wenn es notwendig war, auch die Palpation mit den Fingerkuppen direkt auf den Proc. spinosi an. Die Handhaltung variierte folglich zwischen den Testern und auch zwischen den Probanden.

Die Vorgabe pro Segment nur 2- bis 3-mal zu testen, konnte ebenfalls nicht immer eingehalten werden. Wenn zwei Segmente ähnlich schlecht bewegten oder der Tester nicht sicher war, welches Segment nun das hypomobilste sei, wurde die Bewegung noch einmal wiederholt. Teilweise hatten die Tester das Gefühl, dass die Probanden anfänglich noch etwas angespannter waren und deshalb die Palpation erschwert war. Bei einigen Probanden wurde die Bewegung bis zu 5-mal pro Segment wiederholt.

Das ganze Setting war für die Probanden ungewohnt und manche hatten Schwierigkeiten, auf Kommando locker zu lassen. Das führte dazu, dass die durch den Tester induzierte Bewegung schwieriger durchführbar oder in einem Fall auch „zahnradartig“ war. Es wurde vermutet, dass ein Teil der Patienten mehr Zeit zum „Einstimmen“ gebraucht hätte.

Die angebrachten Markierungen mit dem Kugelschreiber verwischten zum Teil etwas und waren im zweiten Durchgang teilweise schon schlecht erkennbar.

INTERPRETATION:

Interpretationsschwierigkeiten betrafen grundsätzliche Fragen zur Mobilitätstestung: *Wo ist die Grenze zum Normalen? Ist das jetzt noch normal oder schon ein „bisschen“ hypomobil? Ab wann ist eine Hypomobilität erwähnenswert oder vernachlässigbar?*

Außerdem war nicht klar, wie verfahren werden sollte, wenn zwei Segmente gleich hypomobil waren. Da sich die Tester entscheiden mussten, stellt das eine mögliche Fehlerquelle für die Reliabilität dar.

Erschwert wurde die Interpretation dann auch noch durch paradoxes Verhalten der Dornfortsätze. Als Beispiel wurde angegeben, dass bei der Palpation der Lateralflexion anstatt des erwarteten Wegbewegens des Dornfortsatzes ein Entgegenkommen gespürt wurde. Dies wurde als Rotation des Wirbels interpretiert. Ein Tester interpretierte diese Situation nicht als hypomobiles Segment, da ja eine Bewegung spürbar war. Der andere Tester interpretierte es allerdings schon als hypomobiles Segment mit dem Argument, dass, nachdem der Wirbel eine Rotation machte, die Lateralflexion offensichtlich nicht möglich war.

4.8 Datenaufbereitung und Analyse

Die statistische Datenanalyse wurde von Dr. Gebhard Woisetschläger vorgenommen. Dazu wurde die Statistiksoftware „R (Version 3.0.2)“ (R Core Team, 2013) verwendet.

4.8.1 Darstellung der Ergebnisse

Die Darstellung des Datenmaterials bei zwei Untersuchern erfolgt in Form einer Kreuztabelle (siehe Tabelle 5), wie von Bortz und Lienert (2003) empfohlen. Die Ziffern L1/2 bis L5/S1 beschreiben jeweils die Höhe des Segments. Die Ziffer 0 steht für „keine Hypomobilität“ in der Lendenwirbelsäule.

Tabelle 5. Kreuztabelle

		Tester 2						Σ
		Wirbel	0	L1/2	L2/3	L3/4	L4/5	
Tester 1	0	L ₀₀	L ₀₁	L ₀₂	L ₀₃	L ₀₄	L ₀₅	C ₀
	L1/2	L ₁₀	L ₁₁	L ₁₂	L ₁₃	L ₁₄	L ₁₅	C ₁
	L2/3	L ₂₀	L ₂₁	L ₂₂	L ₂₃	L ₂₄	L ₂₅	C ₂
	L3/4	L ₃₀	L ₃₁	L ₃₂	L ₃₃	L ₃₄	L ₃₅	C ₃
	L4/5	L ₄₀	L ₄₁	L ₄₂	L ₄₃	L ₄₄	L ₄₅	C ₄
	L5/S1	L ₅₀	L ₅₁	L ₅₂	L ₅₃	L ₅₄	L ₅₅	C ₅
Σ		F ₀	F ₁	F ₂	F ₃	F ₄	F ₅	n

In der Hauptdiagonale (hier dunkelgrün) werden die übereinstimmenden Ergebnisse dargestellt. Die Summe dieser Diagonale ($L_{00} + L_{11} + L_{22} + L_{33} + L_{44} + L_{55}$) ergibt die Anzahl der gesamt übereinstimmenden Ergebnisse. Die nicht übereinstimmenden Häufigkeiten sind hier zur Veranschaulichung hellgrün hinterlegt. Die Zeilensummen (C) und die Spaltensummen (F), welche jeweils die Summe der Probanden ergeben (n), sind hellblau dargestellt.

4.8.2 Berechnung und Beurteilung von Cohens Kappa

Die Berechnung der Interrater-Reliabilität erfolgte nach Cohens Kappa für zwei Untersucher in den drei Ebenen: Sagittalebene, Frontalebene und Transversalebene. Des Weiteren wurden die Kappa-Werte für die einzelnen Bewegungsrichtungen berechnet: Flexion, Extension, Lateralflexion links, Lateralflexion rechts, Rotation links und Rotation rechts.

Cohens Kappa für zwei Untersucher wird folgendermaßen berechnet (Quelle: Sim & Wright, 2005):

$$k = \frac{\text{gefundene Übereinstimmung} - \text{zufällige Übereinstimmung}}{1 - \text{zufällige Übereinstimmung}}$$

Als Formel dargestellt:

$$k = \frac{P_o - P_c}{1 - P_c}$$

Es gibt verschiedene Interpretationsmöglichkeiten von Cohens Kappa. Am gängigsten und auch von Patijn und dem FIMM Scientific Committee (2004) empfohlen ist die Beurteilung nach Landis und Koch (1977; zitiert nach Sim & Wright, 2005, S. 264), welche in Abbildung 3 aufgelistet ist. Zusätzlich wird bei Kappa-Werten zwischen 0,40 und 0,60 die Berechnung der Signifikanz empfohlen.

$k \leq 0$	=	schlechte Übereinstimmung (poor)
0,01 - 0,20	=	etwas Übereinstimmung (slight)
0,21 - 0,40	=	ausreichende Übereinstimmung (fair)
0,41 - 0,60	=	mittelmäßige Übereinstimmung (moderate)
0,61 - 0,80	=	beachtliche Übereinstimmung (substantial)
0,81 - 1,00	=	fast vollkommene Übereinstimmung (almost perfect)

Abbildung 3. Interpretation von Cohens Kappa nach Landis und Koch (1977; zitiert nach Sim & Wright, 2005)

5 ERGEBNISSE

Im folgenden Kapitel werden die Ergebnisse in Kreuztabellen und mit Kappa-Werten dargestellt (gerundet auf 2 Dezimalstellen), jeweils beginnend mit den zwei Bewegungsrichtungen pro Ebene gefolgt vom Gesamtergebnis pro Ebene. Es konnten die Ergebnisse aller 22 Probanden in die Analyse einbezogen werden.

5.1 Interrater-Reliabilität in der Sagittalebene

5.1.1 Flexion

Aus Tabelle 6 wird ersichtlich, dass bei der Bewegungsprüfung in Flexion drei Segmente übereinstimmend als hypomobil befunden wurden. In zwei Fällen waren sich die Tester einig, dass keine Hypomobilität der LWS in Flexion vorliegt. Aus den fünf übereinstimmenden Ergebnissen ergibt sich ein Kappa-Wert von $k = 0,13$. Tester 2 fand in 91% ($n=20$) der Untersuchungen eine Hypomobilität, hingegen Tester 1 nur in 27% ($n=6$).

Tabelle 6. Häufigkeitsverteilung in Flexion

Segment		Tester 2						Σ
		0	L1/2	L2/3	L3/4	L4/5	L5/S1	
Tester 1	0	2	4	3	1	2	4	16
	L1/2	-	1	-	-	-	-	1
	L2/3	-	-	-	-	-	-	0
	L3/4	-	-	-	1	1	-	2
	L4/5	-	1	-	-	1	1	3
	L5/S1	-	-	-	-	-	-	0
	Σ	2	6	3	2	4	5	22

5.1.2 Extension

Tabelle 7. Häufigkeitsverteilung in Extension

Segment		Tester 2						Σ
		0	L1/2	L2/3	L3/4	L4/5	L5/S1	
Tester 1	0	2	7	1	-	-	2	12
	L1/2	-	-	-	-	-	-	0
	L2/3	-	1	1	-	-	-	2
	L3/4	1	1	1	1	-	-	4
	L4/5	1	1	-	-	1	-	3
	L5/S1	1	-	-	-	-	-	1
	Σ	5	10	3	1	1	2	22

Die Übereinstimmung der segmentalen Bewegungsprüfung in Extension ergibt $k = 0,09$. Drei Segmente sind übereinstimmend hypomobil. In zwei Fällen wurde einstimmig die Beweglichkeit der LWS als normal bewertet. In 46% ($n=10$) fand Tester 1 Bewegungseinschränkungen, Tester 2 in 77% ($n=17$) der Fälle, wie aus Tabelle 7 hervorgeht.

5.1.3 Flexion und Extension gesamt

Für die Flexion und Extension ergibt sich somit eine Übereinstimmung in zehn aus 44 Fällen, $k = 0,10$ (siehe Tabelle 8). Auffällig ist, dass Tester 1 in 64% ($n=28$) der Fälle keine Hypomobilität fand und Tester 2 nur in 16% ($n=7$). Tester 2 stellte bei 12 Probanden gehäuft Bewegungseinschränkungen im Segment L1/2 ($n=16$), v.a. in Extension (siehe Tabelle 7, $n=10$), fest.

In zwei Fällen fanden die Tester eine Hypomobilität im Segment L3/4 bzw. L4/5, jedoch in entgegengesetzten Richtungen (siehe Rohdaten im Anhang, Proband 3 und 16). Diese wurden bei der statistischen Auswertung außer Acht gelassen, da für die Interrater-Reliabilität aus osteopathischer Sicht auch die Richtung der Dysfunktion ausschlaggebend ist.

Tabelle 8. Häufigkeitsverteilung in der Sagittalebene

		Tester 2						Σ
		0	L1/2	L2/3	L3/4	L4/5	L5/S1	
Tester 1	0	4	11	4	1	2	6	28
	L1/2	-	1	-	-	-	-	1
	L2/3	-	1	1	-	-	-	2
	L3/4	1	1	1	2	1	-	6
	L4/5	1	2	-	-	2	1	6
	L5/S1	1	-	-	-	-	-	1
	Σ	7	16	6	3	5	7	44

5.1.4 Übersicht Sagittalebene

Tabelle 9. Übersicht Interrater-Reliabilität Sagittalebene

Übereinstimmung	hypomobil	nicht hypomobil	gesamt	erwartete Übereinst.	k
Flexion	3	2	5	2	0,13
Extension	3	2	5	3	0,09
Sagittalebene	6	4	10	6	0,10

In Tabelle 9 werden die Ergebnisse der Sagittalebene gesammelt dargestellt. Die Interrater-Reliabilität für die segmentale passive Mobilitätstestung in der Sagittalebene ist sehr schwach

und liegt zwischen $k = 0,09 - 0,13$. In den Bewegungsrichtungen Flexion und Extension der Sagittalebene kam es bei 22 Probanden zu je fünf Übereinstimmungen. Davon entsprechen sechs Übereinstimmungen der Zufallserwartung.

5.2 Interrater-Reliabilität in der Frontalebene

5.2.1 Lateralflexion links

In Tabelle 10 werden die Ergebnisse der Lateralflexion links dargestellt. In nur einem Fall wurde eine Übereinstimmung am Segment L3/4 gefunden. In vier Fällen waren sich die Tester bezüglich der normalen Beweglichkeit der Lendenwirbelsäule einig. Die Interrater-Reliabilität beträgt hier $k = 0,04$. Tester 2 fand in 77% ($n=17$) der Fälle eine Hypomobilität, Tester 1 in 36% ($n=8$).

Tabelle 10. Häufigkeitsverteilung Lateralflexion links

		Tester 2						Σ
		0	L1/2	L2/3	L3/4	L4/5	L5/S1	
Tester 1	Segment							
	0	4	3	-	2	3	2	14
	L1/2	-	-	-	-	-	-	0
	L2/3	-	-	-	-	1	2	3
	L3/4	-	-	-	1	1	-	2
	L4/5	1	-	-	1	-	-	2
	L5/S1	-	-	-	-	1	-	1
Σ	5	3	0	4	6	4	22	

5.2.2 Lateralflexion rechts

Aus folgender Tabelle 11 geht hervor, dass sich vier einstimmige Bewegungseinschränkungen in der Lateralflexion rechts finden. In einem Fall waren sich die Therapeuten einig, dass keine Hypomobilität besteht. Der Kappa-Wert beträgt $k = 0,09$. Tester 1 fand in 50% ($n=11$) der Fälle eine Hypomobilität, Tester 2 in 86% ($n=19$).

Tabelle 11. Häufigkeitsverteilung Lateralflexion rechts

		Tester 2						Σ
		0	L1/2	L2/3	L3/4	L4/5	L5/S1	
Tester 1	Segment							
	0	1	4	1	-	2	3	11
	L1/2	1	1	-	-	-	-	2
	L2/3	-	-	1	-	-	-	1
	L3/4	-	2	1	-	-	-	3
	L4/5	1	-	-	2	1	-	4
	L5/S1	-	-	-	-	-	1	1
Σ	3	7	3	2	3	4	22	

5.2.3 Lateralflexion gesamt

In der Frontalebene kam es insgesamt zu fünf Übereinstimmungen bei hypomobilen Segmenten und zu fünf Übereinstimmungen einer frei beweglichen LWS (siehe Tabelle 12). Der Kappa-Wert beträgt $k = 0,07$. Tester 1 fand in 43% ($n=19$) der Untersuchungen eine Hypomobilität, Tester 2 in 81% ($n=36$).

In drei Fällen fanden die Tester Bewegungseinschränkungen zwischen den Segmenten L3/L4 bzw. L4/L5, jedoch jeweils in die entgegengesetzte Bewegungsrichtung. Diese wurden bei der Berechnung der Kappa-Werte nicht einbezogen.

Tabelle 12. Häufigkeitsverteilung Frontalebene

Segment		Tester 2						Σ
		0	L1/2	L2/3	L3/4	L4/5	L5/S1	
Tester 1	0	5	7	1	2	5	5	25
	L1/2	1	1	-	-	-	-	2
	L2/3	-	-	1	-	1	2	4
	L3/4	-	2	1	1	1	-	5
	L4/5	2	-	-	3	1	-	6
	L5/S1	-	-	-	-	1	1	2
Σ		8	10	3	6	9	8	44

5.2.4 Übersicht Frontalebene

Bei der passiven segmentalen Bewegungsprüfung der Lendenwirbelsäule wurden für die Bewegungsrichtungen der Frontalebene insgesamt 10 Übereinstimmungen aus 44 Fällen erreicht, wie in Tabelle 13 übersichtlich dargestellt ist. Davon entsprachen sieben Übereinstimmungen der erwarteten Zufallsübereinstimmung. Die Interrater-Reliabilität für die Untersuchung in Lateralflexion ist sehr niedrig und liegt bei $k = 0,04 - 0,09$.

Tabelle 13. Übersicht Interrater-Reliabilität Frontalebene

Übereinstimmung	hypomobil	nicht hypomobil	gesamt	erwartete Übereinst.	k
Lateralflexion links	1	4	5	4	0,04
Lateralflexion rechts	4	1	5	3	0,09
Frontalebene	5	5	10	7	0,07

5.3 Interrater-Reliabilität in der Transversalebene

5.3.1 Rotation links

Zwei einstimmige Bewegungseinschränkungen fanden die Tester bei der Rotation links (siehe Tabelle 14). In einem Fall waren sie sich einig, dass keine Hypomobilität der Lendenwirbelsäule besteht. Dies entspricht einem Kappa-Wert von $k = 0,04$. Tester 1 fand in 50% ($n=11$) der Fälle keine Hypomobilität der Lendenwirbelsäule, Tester 2 in 4,5% ($n=1$).

Tabelle 14. Häufigkeitsverteilung Rotation links

Segment		Tester 2						Σ
		0	L1/2	L2/3	L3/4	L4/5	L5/S1	
Tester 1	0	1	3	1	1	3	2	11
	L1/2	-	-	-	-	-	-	0
	L2/3	-	3	1	-	1	1	6
	L3/4	-	-	1	-	-	-	1
	L4/5	-	1	-	1	1	-	3
	L5/S1	-	-	-	1	-	-	1
	Σ	1	7	3	3	5	3	22

5.3.2 Rotation rechts

Bei der Mobilitätstestung der Lendenwirbelsäule in Rotation rechts ergaben sich für die Interrater-Reliabilität Kappa-Werte von $k = 0,09$. An vier Segmenten wurde einheitlich eine Bewegungseinschränkung gefunden. In einem Fall waren sich die Tester einig, dass keine Hypomobilität vorliegt (siehe Tabelle 15). Tester 1 fand in 23% ($n=5$) der Fälle keine Einschränkung, Tester 2 in 14% ($n=3$).

Tabelle 15. Häufigkeitsverteilung Rotation rechts

Segment		Tester 2						Σ
		0	L1/2	L2/3	L3/4	L4/5	L5/S1	
Tester 1	0	1	1	1	-	1	1	5
	L1/2	1	1	-	-	-	-	2
	L2/3	-	5	-	-	-	2	7
	L3/4	1	1	1	1	-	-	4
	L4/5	-	-	2	-	2	-	4
	L5/S1	-	-	-	-	-	-	0
	Σ	3	8	4	1	3	3	22

5.3.3 Rotation gesamt

Betrachtet man die Gesamttabelle beider Rotationsrichtungen, gehen daraus sechs einstimmig eingeschränkte Bewegungssegmente für die Rotation links oder rechts hervor (siehe Tabelle 16). In zwei Fällen besteht ein Konsens, dass keine Einschränkung der Lendenwirbelsäule vorliegt. Wie aus der Tabelle der gesammelten Ergebnisse hervorgeht, betrifft dies denselben Probanden (siehe Rohdaten im Anhang, Proband 1). Nach Cohens Kappa entspricht das $k = 0,05$. Tester 1 fand in 9% (n=4) der Fälle keine Hypomobilität, Tester 2 in 36% (n=16).

Tabelle 16. Häufigkeitsverteilung Transversalebene

Segment		Tester 2						Σ
		0	L1/2	L2/3	L3/4	L4/5	L5/S1	
Tester 1	0	2	4	2	1	4	3	16
	L1/2	1	1	-	-	-	-	2
	L2/3	-	8	1	-	1	3	13
	L3/4	1	1	2	1	-	-	5
	L4/5	-	1	2	1	3	-	7
	L5/S1	-	-	-	1	-	-	1
Σ		4	15	7	4	8	6	44

5.3.4 Übersicht Transversalebene

In Tabelle 17 werden die Ergebnisse der Mobilitätstestung in der Transversalebene noch einmal gesammelt dargestellt. Insgesamt wurden acht übereinstimmende Ergebnisse erreicht, wovon allein sechs Fälle der Zufallserwartung entsprachen. Die Kappa-Werte sind auch hier sehr niedrig und liegen unter $k = 0,10$.

Tabelle 17. Übersicht Interrater-Reliabilität Transversalebene

Übereinstimmung	hypomobil	nicht hypomobil	gesamt	erwartete Übereinst.	k
Rotation links	2	1	3	2	0,04
Rotation rechts	4	1	5	3	0,09
Transversalebene	6	2	8	6	0,05

5.4 Zusammenfassende Darstellung der Ergebnisse

In Tabelle 18 wird die Interrater-Reliabilität der sechs physiologischen Bewegungsrichtungen übersichtlich dargestellt. Der Großteil der Kappa-Werte liegt unter $k = 0,10$. Einzige Ausnahme ist die Flexion, in welcher der höchste Wert mit $k = 0,13$ erreicht wurde. Der geringste Wert wurde in der Lateralflexion links bzw. Rotation links mit $k = 0,04$ berechnet. In der Beurteilung nach Landis & Koch spricht man hier in allen Fällen von „etwas Übereinstimmung (slight)“ ($k = 0,01 - 0,20$; s.S.37).

Tabelle 18. Kappa-Werte der sechs Bewegungsrichtungen

Bewegungsrichtung	k
Flexion	0,13
Extension	0,09
Lateralflexion links	0,04
Lateralflexion rechts	0,09
Rotation links	0,04
Rotation rechts	0,09

Fasst man diese Ergebnisse auf den drei Ebenen zusammen, ergibt das für die Transversalebene $k = 0,05$, für die Frontalebene $k = 0,07$ und für die Sagittalebene $k = 0,10$ (siehe Tabelle 19). Auch hier liegt nach Landis & Koch nur „etwas Übereinstimmung“ in der Interrater-Reliabilität vor.

Tabelle 19. Kappa-Werte der drei Ebenen

Bewegungsebene	k
Sagittalebene	0,10
Frontalebene	0,07
Transversalebene	0,05

Vergleicht man die Häufigkeitstabellen miteinander, wird ersichtlich, dass die Anzahl der gefundenen Bewegungseinschränkungen zwischen den Testern, unabhängig von den Übereinstimmungen, sehr divergiert. Dies soll in Tabelle 20 veranschaulicht werden. Durchschnittlich fand Tester 2 in $85 \pm 5\%$ der Fälle eine Bewegungseinschränkung, hingegen Tester 1 nur in $48 \pm 11\%$ der Fälle. Die größte Divergenz ist bei der Flexion ersichtlich, in welcher Tester 1 in 27% ($n=6$) der Untersuchungen eine Dysfunktion fand und Tester 2 in 91% ($n=20$). Der ge-

ringste Unterschied findet sich bei der Rotation rechts, in welcher Tester 1 bei 17 Probanden (77%) eine Bewegungseinschränkung fand und Tester 2 bei 19 Probanden (86%).

Tabelle 20. Häufigkeitsverteilung gefundener Dysfunktionen zwischen den Testern. T1=Tester1; T2=Tester2.

	Dysfunktion		Keine Dysfunktion	
	T1	T2	T1	T2
Flexion	6 (27%)	20 (91%)	16 (72%)	2 (9%)
Extension	10 (45%)	17 (77%)	12 (55%)	5 (23%)
Lateralflexion links	8 (36%)	17 (77%)	14 (64%)	5 (23%)
Lateralflexion rechts	11 (50%)	19 (86%)	11 (50%)	3 (14%)
Rotation links	11 (50%)	21 (95%)	11 (50%)	1 (5%)
Rotation rechts	17 (77%)	19 (86%)	5 (23%)	3 (14%)
Ø Arithm. Mittel ± SD	10,5 ± 2,5 (48 ± 11%)	18,8 ± 1,2 (85 ± 5%)	11,5 ± 2,5 (52 ± 11%)	3,2 ± 1,2 (15 ± 5%)

6 DISKUSSION

6.1 Vergleich der Ergebnisse mit der Literatur

Die bestehende Literatur zur Mobilitätstestung der LWS wurde gewissenhaft recherchiert, stellt jedoch keinen Anspruch auf Vollständigkeit. Um den Rahmen dieser Arbeit nicht zu sprengen und eine bestmögliche „Vergleichbarkeit“ der Studien zu gewährleisten, wurden hier nur Studien beachtet, welche die Interrater-Reliabilität der PPIVMs der Lendenwirbelsäule betreffen. Studien der Halswirbelsäule, Brustwirbelsäule oder der Illiosakralgelenke wurden nicht einbezogen. Studien, welche die Interrater-Reliabilität von PAs oder Hypermobilitäten untersuchten, sind ebenfalls nicht inbegriffen.

Bei der passiven physiologischen Mobilitätstestung der Lendenwirbelsäule im Sitz sind die Ergebnisse der Interrater-Reliabilitätshebung der vorliegenden Studie nach der Interpretation nach Landis und Koch (1977; zitiert nach Sim & Wright, 2005) „etwas übereinstimmend“, mit Kappa-Werten zwischen 0,04 und 0,13. Diese Werte spiegeln die schlechten Ergebnisse früherer Studien wider. Mootz et al. (1989) erreichten Kappa-Werte von $k = -0,17$ bis 0,17 und Keating et al. (1990) Werte von $k = -0,18$ bis 0,31, wobei die Vergleichbarkeit dieser Studien aufgrund methodologischer Mängel, wie in Kapitel 2.4.3 beschrieben, in Frage gestellt wird. Jüngere Studien von Johansson (2006) und Qvistgaard et al. (2007) wiesen schon ein besseres Studiendesign auf, aber auch hier brachten die Ergebnisse keine gute Interrater-Reliabilität und lagen zum Teil unter der Zufallswahrscheinlichkeit. Zum übersichtlicheren Vergleich werden relevante Studien in Tabelle 21 zusammengefasst dargestellt.

Tabelle 21. Interrater-Reliabilitätsstudien der LWS. Daten entnommen aus: Haneline et al., 2008, 2009; Johansson, 2006; Qvistgaard et al., 2007

Autor	Region	Probanden	Ergebnisse
Mootz et al. (1989)	L1-S1	60	$k = -0,17$ bis 0,17
Keating et al. (1990)	T12-S1	46	$k = -0,18$ bis 0,31
Phillips & Twomey (1996)	L1-5	72	$k_w = -0,11$ bis 0,32
Strender et al. (1997)	L5-S1	71	$k = -0,08$ bis 0,75
Lundberg & Gerdle (1999)	Th10-S1	150	$k_w = 0,59$ bis 0,75
Johansson (2006)	L1-L5	20	Flexion: $k = -0,18$ bis 0,21 Extension: $k = -0,25$ bis 0,53
Qvistgaard et al. (2007)	Th12-L1	43	Extension: $k = 0,23$.

6.2 Limitierende Faktoren der Methodologie

Unterschiedlichste Studiendesigns kommen bei der Erhebung der Interrater-Reliabilität zur Mobilitätstestung der Wirbelsäule zum Einsatz. Sicherlich hat jedes davon seine Vor- und Nachteile, aber unweigerlich schleichen sich auch Fehler in jedes Studiendesign ein. Um die Interpretation der vorliegenden Ergebnisse in entsprechende Relation zu setzen, wird in den folgenden Kapiteln die Methodik der vorliegenden Studie nochmal kritisch hinterfragt bzw. nach Verbesserungsansätzen für Folgestudien gesucht.

6.2.1 Studiendesign

Die Standardisierung der Untersuchung, wie sie von Patijn und dem FIMM Scientific Committee (2004) empfohlen wird, wurde in vorliegender Studie geplant. Bei der Durchführung der Untersuchung war dies jedoch nicht immer umsetzbar. Im Feedback gaben die Tester an, dass mit vorgegebener Handhaltung bei der Testung teilweise zu wenig Information gewonnen werden konnte, sodass bei Bedarf variable Handhaltungen zum Einsatz kamen. Von Probanden kam das Feedback, dass sich die Untersuchungen teilweise sehr unterschiedlich anfühlten. Trotz Standardisierung von Untersuchung und Handhaltung bleiben immer subjektive Komponenten, wie z.B. das passive Bewegen des Probanden oder mit wie viel Druck die Palpation ausgeführt wird. Das Ziel der Untersuchung der Wirbelsäule ist es relevante Dysfunktionen aufzufinden. Im Praxisalltag ist die genaue Vorgehensweise und Handhaltung geprägt von persönlichen Vorlieben. Somit spiegelt die Standardisierung der Handhaltung ohnehin kaum die Realität wider. Obwohl diese persönlichen Stile beim Untersuchen möglicherweise dazu führen, dass bei der Interrater-Reliabilitäts-Erhebung im schlimmsten Fall unterschiedliche Methoden verglichen werden (siehe Kapitel 2.4.2.2), entsprächen sie mehr dem therapeutischen Alltag und sollten auch für wissenschaftliche Arbeiten in Zukunft in Betracht gezogen werden. Auch die Vorgabe pro Segment nur 2- bis 3-mal zu testen, konnte nicht immer eingehalten werden. Bei einigen Probanden wurde die Bewegung bis zu 5-mal pro Segment durchgeführt. Ein mobilisierender Effekt kann somit nicht ausgeschlossen werden.

Für die vorliegende Studie wurde versucht, die Interpretation der Ergebnisse zu standardisieren. Die Tester waren angewiesen, die Mobilität nach quantitativen Kriterien zu beurteilen und wurden mit folgenden praxisüblichen Fragen in die Testung geschickt: *Welches der untersuchten Segmente ist das hypomobilste?* bzw. *Welches Segment bedarf deshalb einer genaueren Untersuchung?* Während der Palpation aller fünf Segmente war eine Entscheidung anhand einer dichotomen Skala mit den Möglichkeiten „hypomobil/nicht hypomobil“ zu treffen. Die Tester gaben nach der Untersuchung im Feedback auch an, dass diese Frage eine Schwierigkeit

darstellte. Die Tatsache, dass es keine klaren Richtlinien gibt, ab wann ein Segment eine erwähnenswerte Hypomobilität aufweist bzw. was noch „normal“ ist, stellt somit eine Fehlerquelle dar. Landel et al. (2008) gaben schon zu bedenken, dass keine gemeinsame Definition von „normaler Mobilität“ bestehen kann, da zwei Therapeuten immer unterschiedliche Erfahrungen haben (siehe Kapitel 2.3.2). Außerdem war nicht klar, wie verfahren werden sollte, wenn zwei Segmente gleich hypomobil erschienen. Für zukünftige Studien wäre es hilfreich, in diesen Fällen eine Entscheidungsrichtlinie aufzustellen. D.h., wenn sich die Entscheidung um Nuancen zwischen „noch normal“ und „schon hypomobil“ dreht, könnte man im Zweifelsfall das Segment als „normal“ beweglich einstufen.

Erschwert wurde die Interpretation der Mobilität dann auch noch durch „paradoxes“, d.h. unerwartetes, Verhalten der Dornfortsätze. Als Beispiel wurde im Feedback der Tester angegeben, dass bei der Palpation der Lateralflexion anstatt des erwarteten Wegbewegens des Dornfortsatzes ein Entgegenkommen spürbar war. Dies wurde als Rotation des Wirbels interpretiert. Ein Tester deutete diese Situation nicht als hypomobiles Segment, da ja eine Bewegung spürbar war. Der andere Tester legte es allerdings schon als hypomobiles Segment aus, mit dem Argument, dass, wenn der Wirbel eine Rotation macht, die Lateralflexion offensichtlich nicht möglich ist. Am Ende ist das, was der Tester findet und wie er es interpretiert, so gut man auch versucht die Testung abzustimmen, immer geprägt von der subjektiven Wahrnehmung der Tester.

Auch die Benennung des hypomobilsten Segments pro Bewegungsrichtung darf kritisch betrachtet werden. Zum einen stellt die Benennung der Bewegungsrichtung selbst eine Fehlerquelle dar. Als Beispiel möchte ich Proband Nr. 16 dieser Studie anführen (siehe Rohdaten im Anhang): In der Sagittalebene fand Tester 1 eine Dysfunktion von L4/L5 in Extension, während Tester 2 eine Dysfunktion von L4/L5 in Flexion benannte. D.h., grundsätzlich waren sich die Tester über die Höhe der Hypomobilität einig, nicht aber bei der Richtung der Einschränkung. Zum anderen entspricht das Auffinden und Benennen EINER segmentalen Dysfunktion mehr der Praxis und wurde in jüngeren Studien bereits angestrebt (z.B.: Arndt, 2013; Qvistgaard et al., 2007). Qvistgaard et al. (2007) zeigten, dass die Interrater-Reliabilität sich verbessert, wenn man die Benennung zweier benachbarter Wirbelsegmente als Übereinstimmung wertet. Dies wurde in vorliegender Studie nicht in Erwägung gezogen. Bedenkt man allerdings, dass eine eingebrachte Manipulation nie nur auf ein Segment wirkt, sondern immer auf eine Region, wie von Christensen et al. (2002; zitiert nach Haneline & Young, 2009) behauptet wird, so erscheint es sinnvoll, diese Option in Betracht zu ziehen. Dies würde vielleicht auch erklären warum durch die Manipulation der Wirbelsäule lumbale Schmerzen verringert werden können

(Licciardone et al., 2005), obwohl die Interrater-Reliabilität beim Auffinden von Dysfunktion so schlecht ist.

Als weiteren limitierenden Faktor könnte man die fehlende globale Bewegungsprüfung der Wirbelsäule erwähnen. In der Regel wird in der Praxis die Lendenwirbelsäule dann spezifisch untersucht, wenn bei der globalen Bewegungsprüfung eine Bewegungseinschränkung ersichtlich ist und man eine Dysfunktion eines Segments erwarten kann. Dies wurde bei den teilnehmenden Probanden nicht gemacht, und es kann nicht ausgeschlossen werden, dass möglicherweise nur wenige bis gar keine Dysfunktionen vorhanden waren. Die Interpretation der Ergebnisse wird dadurch erschwert. Arndt (2013) löste diese Frage, indem in ihrer Studie nur Probanden mit einem eingesteiften Segment der Lendenwirbelsäule teilnahmen und somit eine definitive Dysfunktion vorhanden war.

Das Einzeichnen der Markierungen an den Dornfortsätzen gestaltete sich aufgrund der im Sitz erhöhten Spannung des umliegenden Gewebes schwieriger als ursprünglich angenommen. Deshalb wurde bei Bedarf eine zusätzliche kleine Bewegung über den Rumpf induziert, um durch das Öffnen und Schließen der Procc. spinosi die Zwischenräume aufzufinden. Eine besondere Schwierigkeit stellte die Palpation des Proc. spinosus von L5 dar, v.a. auch bei der Unterscheidung von L5/S1/S2. Die Markierungen selbst waren im zweiten Testdurchgang teilweise schon schlecht sichtbar, da sie verwischten. In vorliegender Studie wurden die Markierungen mit einem Kugelschreiber angebracht. Die Verwendung eines wasserfesten Stiftes erscheint in Zukunft obligat.

Die passive physiologische Mobilitätstestung der Lendenwirbelsäule wurde in vorliegender Studie im Sitz durchgeführt, da diese in der osteopathischen Praxis oft Teil der Befundung ist, und bisherige Reliabilitätsstudien darüber wesentliche Mängel aufwiesen. Im Vergleich zur Seitenlage besteht jedoch im Sitz eine erhöhte neuromuskuläre Spannung des umliegenden Gewebes, wodurch die Palpation möglicherweise erschwert wurde. Im Praxisalltag genügt die Untersuchung in einer Ausgangsposition ohnehin nicht, und Therapieentscheidungen werden normalerweise nicht anhand der Testungen in nur einer Ausgangsposition getroffen. Reliabilitätsstudien, welche in nur einer Ausgangsposition untersuchten, wiesen bisher auch immer eine schlechte Interrater-Reliabilität auf (siehe Kapitel 2.4). Die Untersuchung der Lendenwirbelsäule in verschiedenen Ausgangspositionen entspräche also mehr der täglichen Routine und sollte für zukünftige Studien in Betracht gezogen werden, wie dies z. B. bei Qvistgaard et al. (2007) bereits der Fall war.

6.2.2 Die Untersucher

Obwohl aus statistischer Sicht keine Notwendigkeit von mehr als zwei Testern besteht (siehe Kapitel 2.4.2.2), soll dieser Punkt hier kritisch betrachtet werden. Vergleicht man in vorliegender Studie die gefundenen Ergebnisse der beiden Tester, wird ersichtlich, dass ein Tester deutlich mehr Dysfunktionen fand als der andere (siehe Seite 44). Das könnte darauf hindeuten, dass einer der Tester ein Ausreißer ist und das Ergebnis somit nicht auf die Gesamtheit der Osteopathen übertragbar. Es könnte aber auch aus oben beschriebener Problematik der undefinierten Grenzen zwischen normaler/eingeschränkter Mobilität resultieren (siehe oben). Der Einsatz mehrerer Testerpaare und Probandengruppen könnte die Aussagekraft solcher Studien steigern.

Die Erfahrung der teilnehmenden Tester erscheint der Autorin adäquat und auch die erfolgreiche Blindierung der Tester bezüglich der Ergebnisse des anderen und Symptomatik der Probanden kann in vorliegender Studie durch die teilnehmenden Assistentinnen gewährleistet werden.

6.2.3 Die Probanden

Die teilnehmende Probandenzahl ist mit 22 relativ niedrig und limitiert natürlich die Aussagekraft dieser Studie. Die Probanden wurden außerdem via E-mail und SMS rekrutiert. Günstiger wäre es Patienten einzubinden, welche von sich aus die osteopathische Praxis konsultieren. Damit würden die Patienten repräsentativer für die Praxis und somit die Aussagekraft dieser Studie mehr untermauert. Aus logistischen Gründen war das aber nicht umsetzbar. Da die Tester und Assistentinnen an unterschiedlichen Orten in unterschiedlichen Praxen arbeiteten, war es notwendig, die Untersuchung auf einen Termin festzulegen. Nach Auffassung der Autorin hätten die Patienten, welche die osteopathische Praxis aufsuchen, aber ein Recht auf den nächstmöglichen Behandlungstermin. Diesen bis zum geplanten Untersuchungstag hinauszuzögern, wäre ethisch nicht vertretbar. Ebenso wäre eine osteopathische Behandlung vor der Testdurchführung nicht sinnvoll, da sonst bestehende Dysfunktionen eventuell behoben worden wären und man den Testern die Möglichkeit genommen hätte, diese Dysfunktionen zu finden. Somit erschien die Rekrutierung der Probanden via E-mail und SMS aus logistischen und ethischen Gründen am sinnvollsten.

6.3 Ausblick

Zusammengefasst aus obenstehender Diskussion könnten in zukünftigen Studien folgende Punkte Beachtung finden:

- Die Einführung einer Voruntersuchung, im Sinne einer globalen Bewegungsprüfung der Wirbelsäule, mit dem Ziel Probanden mit/ohne Dysfunktionen vorab zu identifizieren;
- Die Durchführung der Untersuchungen in verschiedenen Ausgangspositionen um eine Annäherung des Studiendesigns an die osteopathische Praxis zu erreichen;
- Die Notwendigkeit einer standardisierten Handhaltung für wissenschaftliche Studien, wenn diese doch kaum dem Praxisalltag entspricht;
- Der Einsatz mehrerer Testerpaare und Probandengruppen, um etwaige Ausreißer identifizieren zu können;
- Die Entwicklung detaillierter Entscheidungsrichtlinien für die Interpretation der Ergebnisse, um Bias aus dieser zu minimieren;
- Die Wertung von zwei benachbarten Wirbeln als „Übereinstimmung“, da Manipulationen möglicherweise ohnehin regional wirken und die Sinnhaftigkeit der exakten Höhenbestimmung der Dysfunktion kritisch betrachtet werden darf;

7 KONKLUSION

Die segmentale passive Mobilitätstestung der Wirbelsäule ist eine gängige manuelle Untersuchungsmethode, um Restriktionen zwischen zwei Wirbeln zu finden. Die Interrater-Reliabilität der passiven Mobilitätstestung wurde bisher in vielen Studien als schlecht bewertet, wobei die Qualität dieser Studien häufig bemängelt wurde.

Ziel dieser Studie war es, die Interrater-Reliabilität der spezifisch segmentalen physiologischen Mobilitätstestung der Lendenwirbelsäule zu untersuchen. Vorab wurden durch die Literaturrecherche die Qualitätskriterien für Interrater-Reliabilitätsstudien zusammengefasst und ein reflektiertes Studiendesign erstellt.

Zwei erfahrene Tester führten die Untersuchung an 22 symptomatischen bzw. asymptomatischen Probanden im Sitz durch. Es wurden sämtliche Bewegungsrichtungen untersucht und pro Bewegungsrichtung das hypomobilste Segment benannt. Die Dokumentation erfolgte durch zwei Assistentinnen, welche auch die Blindierung der Tester gewährleisteten. Die Berechnung der Interrater-Reliabilität erfolgte nach Cohens Kappa für zwei Untersucher in den drei Ebenen: Sagittalebene, Frontalebene und Transversalebene. Des Weiteren wurden die Kappa-Werte für die einzelnen Bewegungsrichtungen berechnet: Flexion, Extension, Lateralflexion links, Lateralflexion rechts, Rotation links und Rotation rechts. Die Kappa-Werte liegen hierfür zwischen $k = 0,04$ und $0,13$. Die Ergebnisse decken sich mit denen früherer Literatur und führen zu dem Schluss, dass die Interrater-Reliabilität der spezifisch segmentalen physiologischen Mobilitätstestung der Lendenwirbelsäule im Sitz schlecht ist.

Die anfängliche Forschungsfrage, ob die Untersuchung der Mobilität der einzelnen Segmente der Lendenwirbelsäule im Sitz sowohl an symptomatischen als auch an asymptomatischen Patienten verlässliche, nachvollziehbare Ergebnisse bringt, kann somit mit „Nein“ beantwortet werden. Auf Basis der Ergebnisse sind die Nullhypothesen dieser Studie gültig:

Nullhypothese 1 (a,b,c):

Die palpatorischen Ergebnisse der segmentalen passiven Mobilitätstestung in den drei BEWEGUNGSEBENEN der Lendenwirbelsäule im Sitz sind abhängig vom Untersucher.

Nullhypothese 2 (a,b,c,d,e,f):

Die palpatorischen Ergebnisse der segmentalen passiven Mobilitätstestung in den verschiedenen BEWEGUNGSRICHTUNGEN der Lendenwirbelsäule im Sitz sind abhängig vom Untersucher.

Die Übertragbarkeit dieser Ergebnisse auf die Gesamtheit der Osteopathen scheint fraglich, da die Ergebnisse möglicherweise durch einen Ausreißer beeinträchtigt wurden. Und auch trotz des Versuchs, die Testung und Beurteilungskriterien möglichst zu standardisieren und zu objektivieren, bleibt am Ende das Ergebnis immer geprägt vom subjektiven Empfinden des Testers.

Eine schlechte Interrater-Reliabilität bedeutet aber für die osteopathische Befunderhebung auch nicht, dass diese Untersuchungsmethode im Sitz völlig ungeeignet ist. Es untermauert lediglich, dass Befunde, welche aus nur einem Test hervorgehen, weiterhin kritisch reflektiert werden müssen. Verschiedene Ausgangspositionen zur Befunderhebung, wie sie in der Praxis üblicherweise angewendet werden, empfehlen sich auch für weitere Studien.

Die Relevanz der exakten Höhenbestimmung der Dysfunktion sollte in Zukunft ohnehin nochmal kritisch überdacht werden. Denn wie erklärt sich eigentlich die hohe Wirksamkeit der Manipulation der Lendenwirbelsäule, wenn doch das Auffinden der Dysfunktion bislang schon so unreliaabel war?

Literaturverzeichnis

- Abbott, J. H., Flynn, T. W., Fritz, J. M., Hing, W. A., Reid, D., & Whitman, J. M. (2009). Manual physical assessment of spinal segmental motion: Intent and validity. *Manual Therapy*, 14(1), 36–44. doi:10.1016/j.math.2007.09.011
- Arndt, K. (2013). *Intertester-Reliabilität einer Kombination mehrerer manualtherapeutischer segmentaler Beweglichkeitstests: In Gegenüberstellung mit einem Test zur Identifikation einer Hypomobilität im Bereich der Lendenwirbelsäule*. Krems: Masterthese. Donau-Universität Krems.
- Bergstrom, E., & Courtis, G. (1986). An inter- and intraexaminer reliability study of motion palpation of the lumbar spine in lateral flexion in the seated position. *European Journal of Chiropractic*, 34, 121–41.
- Binkley, J., Stratford, P. W., & Gill, C. (1995). Interrater Reliability of Lumbar Accessory Motion Mobility Testing. *Physical Therapy*, 75(9), 786–792.
- Bisschop, A., Kingma, I., Bleys, R. L., Paul, C. P., van der Veen, A. J., van Royen, B. J. ., & van Dieën, J. H. (2013). Effects of repetitive movement on range of motion and stiffness around the neutral orientation of the human lumbar spine. *Journal of Biomechanics*, 46(1), 187–191. doi:10.1016/j.jbiomech.2012.10.014
- Bortz, J., & Döring, N. (2006). *Forschungsmethoden und Evaluation: Für Human- und Sozialwissenschaftler*. Heidelberg: Springer.
- Bortz, J., & Lienert, G. A. (2003). *Kurzgefasste Statistik für die klinische Forschung: Leitfaden für die verteilungsfreie Analyse kleiner Stichproben* (2. Auflage). Heidelberg; Berlin u.a.: Springer.
- Christensen, H. W., Vach, W., Vach, K., Manniche, C., Haghfelt, T., Hartvigsen, L., & Høilund-Carlsen, P. F. (2002). Palpation of the upper thoracic spine: An observer reliability

- study. *Journal of Manipulative and Physiological Therapeutics*, 25(5), 285–292.
doi:10.1067/mmt.2002.124424
- Ciranna-Raab, C., Fossum, C., & Och, V. (2010). Diagnostische Grundlagen. In T. Liem & T. K. Dobler (Eds.), *Leitfaden Osteopathie: Parietale Techniken* (3. Auflage). München: Elsevier, Urban & Fischer Verlag.
- Degenhardt, B. F., Johnson, J. C., Snider, K. T., & Snider, E. J. (2010). Maintenance and improvement of interobserver reliability of osteopathic palpatory tests over a 4-month period. *JAOA: Journal of the American Osteopathic Association*, 110(10), 579–586.
- Degenhardt, B. F., Snider, K. T., Snider, E. J., & Johnson, J. C. (2005). Interobserver reliability of osteopathic palpatory diagnostic tests of the lumbar spine: improvements from consensus training. *JAOA: Journal of the American Osteopathic Association*, 105(10), 465–473.
- Deore, M., & May, S. (2012). The inter-rater and intra-rater reliability of passive physiological accessory movement assessment of lumbar spine in novice manual therapists. *Journal of Bodywork and Movement Therapies*, 16(3), 289–293.
doi:10.1016/j.jbmt.2011.12.006
- Fossum, C., Sommerfeld, P., & Vincent, L. (2010). Lendenwirbelsäule - Diagnostik. In T. Liem & T. K. Dobler (Eds.), *Leitfaden Osteopathie: Parietale Techniken* (3. Auflage). München: Elsevier, Urban & Fischer Verlag.
- Gonnella, C., Paris, S. V., & Kutner, M. (1982). Reliability in evaluating passive intervertebral motion. *Physical Therapy*, 62(4), 436–444.
- Haneline, M., Cooperstein, R., Young, M., & Birkeland, K. (2008). Spinal motion palpation: A comparison of studies that assessed intersegmental end feel vs excursion. *Journal of Manipulative and Physiological Therapeutics*, 31(8), 616–626.
doi:10.1016/j.jmpt.2008.09.007

- Haneline, M., Cooperstein, R., Young, M., & Birkeland, K. (2009). An annotated bibliography of spinal motion palpation reliability studies. *The Journal of the Canadian Chiropractic Association*, 53(1), 40–58.
- Haneline, M., & Young, M. (2009). A review of intraexaminer and interexaminer reliability of static spinal palpation: a literature synthesis. *Journal of Manipulative and Physiological Therapeutics*, 32(5), 379–386. doi:10.1016/j.jmpt.2009.04.010
- Hestbaek, L., & Leboeuf-Yde, C. (2000). Are chiropractic tests for the lumbo-pelvic spine reliable and valid? A systematic critical literature review. *Journal of Manipulative and Physiological Therapeutics*, 23(4), 258–275.
- Horneij, E., Hemborg, B., Johnsson, B., & Ekdahl, C. (2002). Clinical tests on impairment level related to low back pain: a study of test reliability. *Journal of Rehabilitation Medicine*, 34(4), 176–182.
- Huijbregts, P. A. (2002). Spinal motion palpation: a review of reliability studies. *Journal of Manual & Manipulative Therapy*, 10(1), 24–39.
- Hussy, W., Schreier, M., & Echterhoff, G. (2009). *Forschungsmethoden in Psychologie und Sozialwissenschaften - Für Bachelor*. Berlin, Heidelberg, New York: Springer DE.
- Inscoc, E., Witt, P. L., Gross, M. T., & Mitchell, R. U. (1995). Reliability in Evaluating Passive Intervertebral Motion of the Lumbar Spine. *Journal of Manual & Manipulative Therapy*, 3(4), 135–143. doi:10.1179/jmt.1995.3.4.135
- Johansson, F. (2006). Interexaminer reliability of lumbar segmental mobility tests. *Manual Therapy*, 11(4), 331–336. doi:10.1016/j.math.2005.06.014
- Keating, J. C., Jr, Bergmann, T. F., Jacobs, G. E., Finer, B. A., & Larson, K. (1990). Interexaminer reliability of eight evaluative dimensions of lumbar segmental abnormality. *Journal of Manipulative and Physiological Therapeutics*, 13(8), 463–470.

- Landel, R., Kulig, K., Fredericson, M., Li, B., & Powers, C. M. (2008). Intertester reliability and validity of motion assessments during lumbar spine accessory motion testing. *Physical Therapy, 88*(1), 43–49. doi:10.2522/ptj.20060179
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174.
- Licciardone, J. C., Brimhall, A. K., & King, L. N. (2005). Osteopathic manipulative treatment for low back pain: a systematic review and meta-analysis of randomized controlled trials. *BMC Musculoskeletal Disorders, 6*, 43. doi:10.1186/1471-2474-6-43
- Licciardone, J. C., & Kearns, C. M. (2012). Somatic dysfunction and its association with chronic low back pain, back-specific functioning, and general health: results from the OSTEO-PATHIC Trial. *JAOA: Journal of the American Osteopathic Association, 112*(7), 420–428.
- Liem, T., & Dobler, T. K. (Hrsg.) (2010). *Leitfaden Osteopathie: Parietale Techniken* (3. Auflage). München: Elsevier, Urban & Fischer Verlag.
- Love, R. M., & Brodeur, R. R. (1987). Inter- and intra-examiner reliability of motion palpation for the thoracolumbar spine. *Journal of Manipulative and Physiological Therapeutics, 10*(1), 1–4.
- Lucas, N., & Bogduk, N. (2011). Diagnostic reliability in osteopathic medicine. *International Journal of Osteopathic Medicine, 14*(2), 43–47. doi:10.1016/j.ijosm.2011.01.001
- Lundberg, G., & Gerdle, B. (1999). The relationships between spinal sagittal configuration, joint mobility, general low back mobility and segmental mobility in female homecare personnel. *Scand J Rehab Med, 31*, 197–206.
- Maher, C., & Adams, R. (1994). Reliability of Pain and Stiffness Assessments in Clinical Manual Lumbar Spine Examination. *Physical Therapy, 74*(9), 801–809.

- Maitland, G. D., Hengeveld, E., & Banks, K. (2005). *Maitland's Vertebral Manipulation* (Auflage: 7.). Edinburgh, London, New York, Oxford, Philadelphia, St. Louis, Sydney, Toronto: Elsevier, Butterworth, Heinemann.
- May, S., Littlewood, C., & Bishop, A. (2006). Reliability of procedures used in the physical examination of non-specific low back pain: A systematic review. *Australian Journal of Physiotherapy*, (52), 91–102.
- Mootz, R. D., Keating, J. C., Jr, Kontz, H. P., Milus, T. B., & Jacobs, G. E. (1989). Intra- and interobserver reliability of passive motion palpation of the lumbar spine. *Journal of Manipulative and Physiological Therapeutics*, 12(6), 440–445.
- Patijn, J., (Hrsg.), & International Federation for Manual/Musculoskeletal Medicine (FIMM) Scientific Committee. (2004). Reproducibility and validity studies of diagnostic procedures in manual/musculoskeletal medicine, Protocol formats, 3rd edition. Abgerufen von http://www.fimm-online.com/pub/en/data/objects/reproducibility_validity.pdf
- Phillips, D. R., & Twomey, L. T. (1996). A comparison of manual diagnosis with a diagnosis established by a uni-level lumbar spinal block procedure. *Manual Therapy*, 1(2), 82–87. doi:10.1054/math.1996.0254
- Qvistgaard, E., Rasmussen, J., Laetgaard, J., Hecksher-Sorensen, S., & Bliddal, H. (2007). Intra-observer and inter-observer agreement of the manual examination of the lumbar spine in chronic low-back pain. *European Spine Journal*, 16(2), 277–282. doi:10.1007/s00586-006-0134-8
- R Core Team. (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing (Version 3.0.2). Vienna, Austria. Abgerufen von <http://www.R-project.org/>
- Schneider, M., Erhard, R., Brach, J., Tellin, W., Imbarlina, F., & Delitto, A. (2008). Spinal palpation for lumbar segmental mobility and pain provocation: an interexaminer reliability

- study. *Journal of Manipulative and Physiological Therapeutics*, 31(6), 465–473.
doi:10.1016/j.jmpt.2008.06.004
- Schomacher, J. (2008). Fähigkeit des spezifischen manuellen Bewegens in einzelnen Wirbelsäulensegmenten. *Manuelle Therapie*, 12(03), 113–124. doi:10.1055/s-2008-1027567
- Seffinger, M. A., Najm, W. I., Mishra, S. I., Adams, A., Dickerson, V. M., Murphy, L. S., & Reinsch, S. (2004). Reliability of spinal palpation for diagnosis of back and neck pain: a systematic review of the literature. *Spine*, 29(19), E413–E425.
- Shaw, K. A., Dougherty, J. J., Treffer, K. D., & Glaros, A. G. (2012). Establishing the content validity of palpatory examination for the assessment of the lumbar spine using ultrasonography: a pilot study. *JAOA: Journal of the American Osteopathic Association*, 112(12), 775–782.
- Sim, J., & Wright, C. C. (2005). The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Physical Therapy*, 85(3), 257–268.
- Snider, K. T., Snider, E. J., Degenhardt, B. F., Johnson, J. C., & Kribs, J. W. (2011). Palpatory accuracy of lumbar spinous processes using multiple bony landmarks. *Journal of Manipulative and Physiological Therapeutics*, 34(5), 306–313.
doi:10.1016/j.jmpt.2011.04.006
- Stochkendahl, M. J., Christensen, H. W., Hartvigsen, J., Vach, W., Haas, M., Hestbaek, L., ... Bronfort, G. (2006). Manual examination of the spine: a systematic critical literature review of reproducibility. *Journal of Manipulative and Physiological Therapeutics*, 29(6), 475–485.e10. doi:10.1016/j.jmpt.2006.06.011
- Strender, L. E., Sjöblom, A., Sundell, K., Ludwig, R., & Taube, A. (1997). Interexaminer reliability in physical examination of patients with low back pain. *Spine*, 22(7), 814–820.
- Triano, J. J., Budgell, B., Bagnulo, A., Roffey, B., Bergmann, T., Cooperstein, R., ... Tepe, R. (2013). Review of methods used by chiropractors to determine the site for applying

manipulation. *Chiropractic & Manual Therapies*, 21(1), 36. doi:10.1186/2045-709X-21-36

Van Trijffel, E., Anderegg, Q., Bossuyt, P. M. M., & Lucas, C. (2005). Inter-examiner reliability of passive assessment of intervertebral motion in the cervical and lumbar spine: a systematic review. *Manual Therapy*, 10(4), 256–269. doi:10.1016/j.math.2005.04.008

Wiesner, R., & Bucher-Dollenz, G. (s.a.). *IMTA (international maitland teachers association) Kurshandbuch, Level 1 - Kurs (Version 9.1.)*. Stuttgart: Georg Thieme Verlag.

Tabellenverzeichnis

<i>Nummer</i>	<i>Seite</i>
<i>Tabelle 1: Interrater-Reliabilitätsstudien der LWS im Sitz</i>	<i>20</i>
<i>Tabelle 2: Interrater-Reliabilitätsstudien der LWS beurteilt nach Bewegungsausmaß</i>	<i>21</i>
<i>Tabelle 3: Demographische Verteilung der Probanden</i>	<i>28</i>
<i>Tabelle 4: Schmerzsituation LWS der Probanden</i>	<i>28</i>
<i>Tabelle 5: Kreuztabelle</i>	<i>36</i>
<i>Tabelle 6: Häufigkeitsverteilung in Flexion</i>	<i>38</i>
<i>Tabelle 7: Häufigkeitsverteilung in Extension</i>	<i>38</i>
<i>Tabelle 8: Häufigkeitsverteilung in der Sagittalebene</i>	<i>39</i>
<i>Tabelle 9: Übersicht Interrater-Reliabilität Sagittalebene</i>	<i>39</i>
<i>Tabelle 10: Häufigkeitsverteilung Lateralflexion links</i>	<i>40</i>
<i>Tabelle 11: Häufigkeitsverteilung Lateralflexion rechts</i>	<i>40</i>
<i>Tabelle 12: Häufigkeitsverteilung Frontalebene</i>	<i>41</i>
<i>Tabelle 13: Übersicht Interrater-Reliabilität Frontalebene</i>	<i>41</i>
<i>Tabelle 14: Häufigkeitsverteilung Rotation links</i>	<i>42</i>
<i>Tabelle 15: Häufigkeitsverteilung Rotation rechts</i>	<i>42</i>
<i>Tabelle 16: Häufigkeitsverteilung Transversalebene</i>	<i>43</i>
<i>Tabelle 17: Übersicht Interrater-Reliabilität Transversalebene</i>	<i>43</i>
<i>Tabelle 18: Kappa-Werte der sechs Bewegungsrichtungen</i>	<i>44</i>
<i>Tabelle 19: Kappa-Werte der drei Ebenen</i>	<i>44</i>
<i>Tabelle 20: Häufigkeitsverteilung gefundener Dysfunktionen zwischen den Testern</i>	<i>45</i>
<i>Tabelle 21: Interrater-Reliabilitätsstudien der LWS</i>	<i>46</i>

Abbildungsverzeichnis

<i>Nummer</i>	<i>Seite</i>
<i>Abbildung 1: Markierung der Processus spinosi</i>	<i>31</i>
<i>Abbildung 2: Testerposition und Handhaltung</i>	<i>33</i>
<i>Abbildung 3: Interpretation von Cohens Kappa nach Landis und Koch</i>	<i>37</i>

Abkürzungsverzeichnis

FIMM	International Federation for Manual/Musculoskeletal Medicine
LWS	Lendenwirbelsäule
PAs	posterior-anterior gerichteter Druck
PAIVMs	passive accessory intervertebral movements
PPIVMs	passive physiological intervertebral movements
Proc. spinosus	Processus spinosus, Dornfortsatz
Procc. spinosi	Processus spinosi (Pl.), Dornfortsätze

Anhang

INFORMATION FÜR STUDIENTEILNEHMER

Masterarbeit: Silvia Maurer

Tel.Nr.: 0650/9216757

e-mail: [REDACTED]

Thema: Interrater Reliabilität der segmentalen passiven Mobilitätstestung der Lendenwirbelsäule im Sitz mit markierten Processi spinosi.

Studienort: Praxisgemeinschaft für Physiotherapie und Osteopathie, Rainerstraße 18a, 4020 Linz

Sehr geehrte Teilnehmerin! Sehr geehrter Teilnehmer!

Vielen Dank für Ihre Teilnahme an meiner Studie.

In der Physiotherapie, Osteopathie und Manualtherapie ist die segmentale (d.h. zwischen den einzelnen Wirbeln) Untersuchung der Lendenwirbelsäule schon lange eine gängige Methode um Bewegungseinschränkungen einzelner Wirbel zu finden. Die Ergebnisse dieser Untersuchung unterstützen in vielen Fällen die Entscheidungsfindung für das weitere Behandlungsprocedere.

Zusätzlich wird in der Schulmedizin als auch in der Komplementärmedizin der Ruf nach evidenzbasierten Untersuchungs- und Behandlungsmethoden immer lauter. Das heißt das viele langjährige Untersuchungs- und Behandlungskonzepte nach und nach durch kritische Studien auf ihre Aussagekraft bzw. Wirksamkeit überprüft werden.

Ziele

Das Ziel meiner Studie ist es herauszufinden, ob verschiedene Therapeuten bei der segmentalen Untersuchung der Lendenwirbelsäule im Sitz auf gleiche Ergebnisse kommen (=Interrater-Reliabilität). Oder ob diese Ergebnisse sehr subjektiv und vom Therapeuten abhängig sind.

Ablauf

Die Untersuchung der Lendenwirbelsäule findet einmalig statt und nimmt in etwa eine Stunde Ihrer Zeit in Anspruch.

Vor der Testung selbst werden die Dornfortsätze (die tastbaren Knochenvorsprünge) an Ihrer Lendenwirbelsäule bestimmt und markiert um sicherzustellen dass die Therapeuten dieselben Segmente untersuchen.

Sie werden nacheinander von zwei Therapeuten in zwei getrennten Räumen untersucht. Dabei sitzen Sie auf einer Liege und die Therapeuten tasten bei den Beweglichkeitstests Ihre Lendenwirbelsäule ab. Um die Testergebnisse nicht zu beeinflussen ist das Sprechen mit den Testern nicht gestattet.

Risiken und Nebenwirkungen

Es bestehen keine Risiken und Nebenwirkungen.

Datenschutz

Ihre Daten werden vertraulich behandelt und in anonymisierter Form abgespeichert.

Auf Wunsch teile ich Ihnen nach Beendigung der Studie gerne das Ergebnis mit.

Muss ich was mitbringen?

Bitte tragen Sie gemütliche Kleidung aus der Sie schnell raus- und reinschlüpfen können.

Vielen herzlichen Dank für Ihre Unterstützung!

Als Entschädigung für den Ihnen entstandenen Aufwand erhalten Sie einen Gutschein für eine Massage oder Craniosacral-Therapie, einzulösen bei der Studienleiterin persönlich. Vor oder nach der Testung können Sie sich jederzeit an einem Kuchenbuffet laben.

Sollten Sie noch Fragen haben, zögern Sie nicht mich telefonisch oder via e-mail zu kontaktieren!

Silvia Maurer

Einverständniserklärung zur Teilnahme an der Studie:

Interrater Reliabilität der segmentalen passiven Mobilitätstestung der Lendenwirbelsäule im Sitz mit markierten Processi spinosi.

Studienleiterin: Silvia Maurer

Studienort: Praxisgemeinschaft für Physiotherapie und Osteopathie, Rainerstr. 18a, 4020 Linz

Name des Teilnehmers: _____

Ich wurde von der unterzeichnenden Studienleiterin mündlich und schriftlich über die Ziele, den Ablauf der Studie, über die zu erwartenden Wirkungen, über mögliche Vor- und Nachteile sowie über eventuelle Risiken informiert.

Ich habe die zur oben genannten Studie abgegebene schriftliche Information gelesen und verstanden. Meine Fragen im Zusammenhang mit der Teilnahme an dieser Studie sind mir zufriedenstellend beantwortet worden. Ich kann die schriftliche Information behalten und erhalte eine Kopie dieser Einverständniserklärung.

Ich hatte genügend Zeit, um die Entscheidung zu treffen, an der Studie teilzunehmen.

Ich weiß, dass meine persönlichen Daten nur in anonymisierter Form gespeichert werden.

Ich nehme an dieser Studie freiwillig teil. Ich kann jederzeit und ohne Angabe von Gründen meine Zustimmung zur Teilnahme widerrufen, ohne dass mir deswegen Nachteile entstehen.

Im Interesse meiner Gesundheit kann mich die Testleiterin oder der Testleiter jederzeit von der Studie ausschließen

Ort, Datum

Unterschrift der Teilnehmerin/des Teilnehmers

Unterschrift der Studienleiterin

Proband Nr. (von Studienleiterin auszufüllen):.....

TEILNEHMERFRAGEBOGEN

Name:.....

Geschlecht: m w

Alter:.....

Größe:cm

Gewicht:kg

RÜCKENSCHMERZEN

1. Haben sie Schmerzen in der Lendenwirbelsäule? JA NEIN

2.a Falls JA - Seit wann: _____

2.b Falls NEIN:

Hatten sie innerhalb der letzten 6 Wochen Schmerzen

in der Lendenwirbelsäule? JA NEIN

Untersuchungsprotokoll

Untersuchungsdatum: 3.5.2014

Untersucher:.....

Assistenz:.....

Proband Nr.:

Kennzeichne das **hypomobilste** Segment!

	L1/L2	L2/L3	L3/L4	L4/L5	L5/S1
FLEXION	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
EXTENSION	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
LAT-FLEX links	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
LAT-FLEX rechts	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ROTATION links	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ROTATION rechts	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Rohdaten, vom 3. Mai 2014

		Sagittalebene				Frontalebene				Transversalebene			
Proband	Ersttester	Flexion		Extension		Lateralflexion links		Lateralflexion rechts		Rotation links		Rotation rechts	
		Tester1	Tester2	Tester1	Tester2	Tester1	Tester2	Tester1	Tester2	Tester1	Tester2	Tester1	Tester2
1	2	0	0	5	0	0	0	1	0	0	0	0	0
2	2	0	2	0	5	2	5	5	5	2	5	1	0
3	2	3	4	3	3	2	4	3	1	3	2	3	0
4	2	1	1	2	2	3	3	2	2	2	1	2	1
5	2	0	1	0	1	0	1	1	1	0	1	3	1
6	2	0	5	0	5	2	5	0	5	0	5	2	5
7	2	0	5	0	0	0	5	3	1	0	5	0	1
8	2	0	2	0	2	0	0	0	2	0	2	0	2
9	2	0	0	0	0	0	0	0	0	2	1	2	1
10	2	0	2	0	1	0	1	0	1	4	1	4	2
11	1	0	1	3	0	0	0	0	1	2	1	2	1
12	2	4	4	4	4	0	5	0	1	4	3	4	4
13	1	4	1	4	1	0	4	0	5	0	4	2	5
14	2	0	4	0	1	0	4	0	4	0	1	1	1
15	1	0	1	0	1	5	4	4	4	5	3	4	2
16	1	0	4	4	0	3	4	0	5	0	4	0	5
17	1	0	3	3	1	4	3	4	3	0	3	3	3
18	1	0	5	0	1	0	3	3	2	2	2	3	2
19	1	3	3	3	2	0	4	4	3	4	4	0	4
20	1	4	5	0	1	0	3	0	1	0	1	2	1
21	1	0	1	2	1	0	1	4	0	2	4	2	1
22	1	0	5	0	1	4	0	0	4	0	4	4	4

Interrater Reliability For Segmental Passive Motion Palpation Of The Lumbar Spine In Sitting Position With Marked Spinous Processes

Silvia Maurer

Danube University Krems – Department for Health Sciences and Biomedicine

ABSTRACT

Objective: Segmental passive motion palpation is a common method to detect restrictions in the spine and frequently serves as a basis for decisions concerning further treatment. The purpose of this study was to test the interrater reliability of the specific segmental physiological motion palpation of the lumbar spine in sitting position.

Methods: Two raters examined the lumbar spine of 22 participants. Some of the participants were symptomatic and some were asymptomatic. The spinous processes of L1 to S1 were marked before the examination. The tests were performed in sitting position; the most hypomobile segment was identified for each direction of movement. The participants were unknown to the raters. Two assistants blinded the raters to the symptoms of the participants and documented the findings. The interrater reliability was calculated by using Cohen's kappa for two raters, for the three planes, and for each of the directions of movement.

Results: The interrater reliability for motion palpation of the lumbar spine was poor, lying within the range of chance agreement. The kappa values ranged from $k = 0.04$ to $k = 0.13$. One of the raters identified a dysfunction in $85 \pm 5\%$ of the cases, whereas the other one identified a dysfunction in only $48 \pm 11\%$ of the cases.

Conclusion: Treatment decisions based on the results of passive motion palpation of the lumbar spine should be critically reflected. However, the general validity for osteopaths of the findings of the study at hand must be questioned, as they are possibly affected by an outlier.

Keywords: mobility, motion palpation, lumbar spine, interrater reliability, segmental examination

INTRODUCTION

Passive motion palpation is a common method used in physiotherapy, osteopathy and other manual therapy concepts to detect restrictions between two vertebrae. Over the last approximately ten years, several narrative and systematic reviews of the reliability of motion palpation of the spine have been published, among them: Haneline, Cooperstein, Young, & Birkeland (2008, 2009); Hestbaek & Leboeuf-Yde (2000); Huijbregts (2002); May, Littlewood, & Bishop (2006); Seffinger et al. (2004); Stochkendahl et al. (2006); Triano et al. (2013); van Trijffel, Anderegg, Bossuyt, & Lucas (2005). The majority of these came to the conclusion that the reliability of passive segmental motion palpation of the spine was poor. Triano et al. and Seffinger et al. also reported conflicting results of interrater reliability. However, due to differing techniques of palpation, statistical analyses and study designs, the comparability of the studies is limited (Seffinger et al., 2004; Haneline et al., 2008). Therefore, in their review, Haneline et al. (2008) distinguished the studies by assessment criteria. Their result was that it did not influence the interrater reliability whether the spinal motion palpation was assessed in terms of quality of motion (end feel) or quantity of movement (excursion of the segments).

According to Snider et al. (2011), the poor reliability in identifying lumbar spinous processes could be a possible reason for poor reliability. Only few studies marked the spinous processes of the lumbar spine before the examination. Seffinger et al. (2004, p. 419) question “whether the palpatory tests are indeed measuring what they are intending to measure”. If they do not, a high reliability of this method is very unlikely.

Many authors criticise above all the poor quality of previous palpation studies (Haneline et al., 2009; May et al., 2006; Seffinger et al., 2004; Stochkendahl et al., 2006; Triano et al., 2013; van Trijffel et al., 2005), especially studies published before 1985 (Hestbaek & Leboeuf-Yde, 2000). In recent years the quality of studies has improved (Seffinger et al., 2004). The criticism of the poor quality of studies concerned the following aspects: participants, raters, study design and description, data preparation and statistical methods.

In general, in the more recent literature, a considerably higher number of studies can be found on passive ACCESSORY intervertebral movements (e.g. Binkley, Stratford & Gill, 1995; Degenhardt, Snider, Snider & Johnson, 2005; Deore & May, 2012; Horneij, Hemborg, Johnsson & Ekdahl, 2002; Maher & Adams, 1994; Schneider et al., 2008; Shaw, Dougherty, Treffer & Glaros, 2012) than on passive PHYSIOLOGICAL intervertebral movements (PPIVMs). Earlier studies which tested the reliability of lumbar PPIVMs in sitting position consistently showed a lack of quality owing to the inclusion of only asymptomatic participants, no marking of the spinous processes or using inappropriate statistical methods (Haneline et al., 2009; Huijbregts,

2002). All these studies were performed before the year 2000. After that only few studies were published examining the PPIVMs on the lumbar spine.

Johansson (2006) assessed the interrater reliability of motion palpation of the lumbar spine with the subjects lying on their side. Three physiotherapists tested the flexion and extension from L1 to L5 on 20 symptomatic and asymptomatic subjects. The results for flexion ranged from $k = -0.18$ to $k = 0.21$ ($k_w = -0.14$ to 0.28); the results for extension ranged from $k = -0.25$ to $k = 0.53$ ($k_w = -0.09$ to 0.56).

Qvistgaard, Rasmussen, Laetgaard, Hecksher-Sorensen and Bliddal (2007) improved the methodology of the study (symptomatic subjects, marked segments, independent observers, blinded raters, statistical calculation using Cohen's kappa). The kappa values for perfect match (meaning "perfect agreement ... was defined as positive result"; p. 279) ranged from $k = 0.21$ to $k = 0.23$; the kappa values for acceptable match (meaning "acceptable agreement ... was defined as positive result on neighbouring segmental levels"; p. 279) ranged from $k = 0.44$ to $k = 0.57$.

It can be subsumed that there are only few studies of good quality assessing the PPIVMs of the lumbar spine and that there is hardly any improvement of reliability with improving quality of the studies.

The aim of the study at hand was to test the interrater reliability of motion palpation of the lumbar spine in sitting position. In a first step, an extensive review of the current literature was conducted in order to identify quality criteria for interrater reliability studies on the spine, which are listed below:

- subjects representing clinical practice (i.e. symptomatic AND asymptomatic participants)
- number of subjects >20
- subjects recruited in a randomised manner
- alternating or randomised order of examination
- raters blinded to subjects' history and findings of the other raters
- controlled by independent observers
- standardisation of test procedure and interpretation of findings
- detailed protocols of the study
- marked spinous processes
- statistical analysis using Cohen's kappa

METHODS

Study design

The study at hand is a methodological study.

The interrater reliability of passive segmental motion palpation of the lumbar spine was tested. For this purpose, symptomatic (low back pain) as well as asymptomatic participants were recruited. Two experienced raters performed the examination of the lumbar spine after marking of the spinous processes by the author of the study. The statistics was calculated using Cohen's kappa.

Raters

Two colleagues of the author volunteered to perform the examination. Both raters were students of the Vienna School of Osteopathy in the 6th year / 3rd term of the course "Osteopathy" at the Danube University Krems.

Rater 1 was 40 years old and had 12 years of professional experience as a physiotherapist. She had been using the segmental examination of the spine regularly since the beginning of her osteopathy training in 2008.

Rater 2 was 42 years old and had 13 years of professional experience as a physiotherapist. He had also been using the segmental examination of the spine since the beginning of his osteopathy training in 2008.

Assistants

Both raters had an assistant. The assistants were a physiotherapist with 4 years of professional experience and a physiotherapist / osteopath (in training) with 8 years of professional experience. Their function was to ensure the blinding of the raters, so that there was no exchange of information between rater / participant and rater / rater. The assistants also documented the findings of the examination.

Participants

Participants with low back pain as well as participants without low back pain were recruited. The minimum age was 18. The following exclusion criteria were applied:

- surgical interventions on trunk, hips or arms in the 6 weeks prior to the study
- inability to sit due to acute pain
- acute disc prolapse with radicular symptoms
- pregnancy

In order to obtain an acceptable sample size for the statistical measurement (>20), 24 subjects had to be recruited. The recruitment was carried out via e-mail and SMS in the circle of friends, relatives and acquaintances, asking them to also forward the request to other people. 26 people volunteered. One of these was excluded because of radicular symptoms, and one was excluded because of pregnancy, resulting in an initial total of 24 test subjects. Before the examination, the subjects received an e-mail informing them about the content and objectives of the study, and they had to sign a letter of consent. As a compensation for the expenditure of time, every participant received a voucher for a 45-minute massage. Two test subjects dropped out, one because of illness and one because the subject forgot the appointment. Consequently, 22 subjects (n=22) finished the study, 12 women and 10 men. The average age was 32.6 ± 5.9 years (min = 24, max = 51, median = 31). The demographic data are illustrated in Table 1.

Table 1. Participant characteristics

Participant characteristics (n=22, f=12, m=10)			
	Age (years)	Height (cm)	Weight (kg)
Mean	32.6	174.8	75.2
SD	5.9	8.7	13.0
Median	31	175	75
Min	24	161	56
Max	51	189	100

8 participants were suffering from low back pain at the moment of examination. 14 participants were not suffering from pain, but 9 of them had suffered from a low back pain within 6 weeks prior to the examination (Table 2). 5 participants were completely free of low back pain, which did not rule out pain in the thoracic or cervical spine.

Table 2. Participants' state of pain in the lumbar spine

Low back pain	n	%
Current pain	8	36%
Within the last 6 weeks	9	41%
No low back pain	5	23%
Overall	22	100%

Blinding

Participant-rater blinding: The raters did not know the participants, their history or their symptoms. To avoid exchange of information, the raters and their participants were instructed not to talk to each other, except for greeting or necessary instructions during the examination. In order to ensure this, the raters were accompanied by an assistant.

Rater-rater blinding: As the raters were not strictly separated (they entered the examination rooms one after the other and met in the corridor), they were instructed not to talk about the participants or their findings. This was also ensured by the assistants.

Examination

Instruction of the raters

Three and a half weeks before the study was carried out, the process of examination was discussed in detail with the raters and assistants and the examination procedure was adjusted. The handhold of the palpating hand was discussed and defined, the assessment criteria were explained and defined, and the form of documentation to be followed was specified.

Process

The examination was performed on 3rd May 2014. Three participants were scheduled for every 30 minutes in random order. They were separated into three rooms in the order of their arrival. First the author of the study entered the room. The participants answered the questionnaire and signed a letter of consent. After that, the author of the study marked the spinous processes of the lumbar spine and S1. When the author left, the first rater entered the room and performed the motion palpation. The second rater did not enter the room until the first rater left. Approximately 10 minutes were allowed for the examination of each participant. A 30-minute break was taken after the first 12 participants. Before the break, the motion palpation was carried out first by rater 2 and then by rater 1 (n=12; 55%). After the break, it was carried out first by rater 1 and then by rater 2 (n=10; 45%).

Marking the spinous processes of the lumbar spine

Due to the poor reliability in identifying the lumbar segments, the spinous processes of the lumbar spine and the first sacral vertebra were marked by the author of the study before the examination (Fig. 1). The 3rd lumbar vertebra was labelled. During the marking, the participants were in sitting position. For a higher accuracy in finding L1-L4, multiple bony landmarks were used, as recommended by Snider et al. (2011).



Fig.1. Marking the spinous processes

Procedure of the passive segmental physiological motion palpation of the lumbar spine

In this study, the raters examined the segments of the lumbar spine in the six directions of movement. They had to identify the most “hypomobile” segment of each direction of movement. Therefore the participants were examined as follows:

Position of participants: The participants were sitting, with their arms crossed in front of their body and their hands placed on their shoulders. They were sitting at the end of the divan bed, so that the rater could change position if necessary.

Position of rater: The rater stood beside the participant. One arm reached from ventral around the upper part of the body and induced the motion (Fig. 2). The other hand lay dorsal and palpated between the spinous processes with the fingertips (Fig. 1). It was the raters’ choice on which side to stand. A change of position during the examination was allowed.

Test and assessment: While the ventral arm induced a flexion, extension, lateral flexion or rotation motion in the lumbar spine, via the upper body, the dorsal hand palpated in between the spinous processes. The opening and closing between the lumbar segments was tested and assessed following quantitative criteria. During the palpation of all five lumbar segments, the raters had to decide between “hypomobile” and “not hypomobile” on a dichotomous scale. The movement was induced a maximum of 2-3 times per segment in order to avoid the effect of mobilisation. After testing each direction of movement, the raters had to decide which level was the most hypomobile. The assistants supervised the process and documented the findings on separate sheets, which were collected in an envelope.



Fig. 2. Rater position and handhold

Data preparation

The statistical data preparation was carried out by Dr. Gebhard Woisetschläger using the statistical software “R (Version 3.0.2)” (R Core Team, 2013). The data for the two raters were presented in a cross tabulation, as recommended by Bortz and Lienert (2003).

RESULTS

The interrater reliability of the six directions of movements is listed in Table 3. Most of the values lie under $k = 0.10$. The only exception was found in the case of flexion, in which the highest value was $k = 0.13$. The lowest values were calculated for lateral flexion left and for rotation left, with $k = 0.04$ in both cases. In the assessment according to Landis and Koch, this means that there was only slight agreement on the dysfunctions identified in each direction of movement (Landis & Koch, 1977; cited in Sim & Wright, 2005).

Table 3. Kappa for the six directions of movement

Direction of movement	k
Flexion	0.13
Extension	0.09
Lateral flexion left	0.04
Lateral flexion right	0.09
Rotation left	0.04
Rotation right	0.09

Summing up these values for the three planes of movement, the results are $k = 0.05$ for the transverse plane, $k = 0.07$ for the frontal plane and $k = 0.10$ for the sagittal plane (Table 4).

Table 4. Kappa for the three planes of movement

Plane of movement	k
Sagittal plane	0.10
Frontal plane	0.07
Transverse plane	0.05

The frequency table for the identified dysfunctions (Table 5) clearly shows that the numbers of dysfunctions identified by each rater diverge significantly. On average, rater 2 identified dysfunctions in $85 \pm 5\%$ of the cases, whereas rater 1 identified dysfunctions in only $48 \pm 11\%$ of the cases. The greatest divergence was in the direction of flexion, in which rater 1 identified a dysfunction in 27% ($n=6$) and rater 2 in 91% ($n=20$). The lowest divergence was in rotation right, in which rater 1 identified a dysfunction in 17 participants (77%) and rater 2 in 19 participants (86%).

Table 5. Frequency of dysfunctions identified by each rater. R1 = rater 1; R2 = rater 2.

	Dysfunction		No dysfunction	
	R1	R2	R1	R2
Flexion	6 (27%)	20 (91%)	16 (72%)	2 (9%)
Extension	10 (45%)	17 (77%)	12 (55%)	5 (23%)
Lateral flexion left	8 (36%)	17 (77%)	14 (64%)	5 (23%)
Lateral flexion right	11 (50%)	19 (86%)	11 (50%)	3 (14%)
Rotation left	11 (50%)	21 (95%)	11 (50%)	1 (5%)
Rotation right	17 (77%)	19 (86%)	5 (23%)	3 (14%)
Ø Mean average \pm SD	10.5 ± 2.5 ($48 \pm 11\%$)	18.8 ± 1.2 ($85 \pm 5\%$)	11.5 ± 2.5 ($52 \pm 11\%$)	3.2 ± 1.2 ($15 \pm 5\%$)

DISCUSSION

Comparing the results with the literature

For the passive physiological motion palpation of the lumbar spine in sitting position, the results for the interrater reliability of this study reflect only slight agreement (assessed according to Landis and Koch, 1977; cited in Sim & Wright, 2005), with kappa values between 0.04 and 0.13. This result confirms the results of earlier studies. Mootz, Keating, Kontz, Milus and Jacobs (1989) reached kappa values from $k = -0.17$ to $k = 0.17$ and Keating, Bergmann, Jacobs, Finer and Larson (1990) reached kappa values from $k = -0.18$ to $k = 0.31$. However, due to the poor quality of these studies, their comparability may be questioned. More recent studies by Johansson (2006) and Qvistgaard et al. (2007) showed improved study designs, but also here the results showed poor interrater reliability and laid partly under the range of chance agreement.

Limiting factors of the methodology

Study design

The standardisation of the examination, as recommended by Patijn and the FIMM Scientific Committee (2004), was chosen for this study. However, it was not always practical during the examination. In their feedback, the raters pointed out that the information that could be gathered by applying the indicated handhold was sometimes insufficient, so that in some cases other handholds were used. They also pointed out significant differences in certain aspects of the examination. Despite standardisation of examination and handhold, subjective components will always remain, for instance the passive movement of the participant or the intensity of the pressure applied in palpation. Furthermore, the instruction to test every segment only 2-3 times could not always be followed; in some participants the motion was carried out up to 5 times per segment. Therefore a mobilising effect cannot be ruled out. To make further studies more practical, it could be useful to refrain from standardising the handhold.

For the study at hand it was also attempted to standardise the interpretation of the findings. The raters were instructed to assess the motion following quantitative criteria. During the palpation of all five lumbar segments a dichotomous decision between “hypomobile” and “not hypomobile” had to be made and the most hypomobile segment had to be named. In some cases this was difficult. The fact that no clear definitions exist as to what is considered “normal” and what is considered “hypomobility” could be a source of error. For further studies such definitions could be very useful, establishing, for instance, that in the case of doubt the segment will

be assessed as “normal”. Furthermore it was not clear how to proceed with two segments which seemed equally hypomobile or showed “paradox”, i.e. unexpected, movements of the spinous processes. For example, the raters mentioned in their feedback that while palpating the lateral flexion, instead of the expected movement of the spinous process away from them (the raters), a movement towards them was felt. This was interpreted as rotation of the vertebra. One rater assessed this situation as hypomobile, the other did not. Eventually everything that a rater finds and how it is interpreted, irrespectively of the degree of standardisation, is always biased by the subjective perception of the rater.

Another aspect that should be critically contemplated is the direction of the dysfunction. On the one hand, the naming of the direction of the dysfunction could be a source of error. For example, in the case of participant no. 16, both raters identified a dysfunction of L4/L5 in the sagittal plane (see annex), but rater 1 identified the dysfunction in the extension, while rater 2 identified a hypomobility in the flexion. This means that both raters agreed on the location of the dysfunction, but not on the direction.

On the other hand, identifying and naming only ONE segmental dysfunction is closer to real-life practice and was already the aim of more recent studies (e.g. Arndt, 2013; Qvistgaard et al., 2007). Qvistgaard et al. (2007) also showed that the interrater reliability improves when naming two adjacent segments counts as agreement (see introduction). This was not considered in the study at hand. But bearing in mind that a manipulation always affects a region, and not only a segment, as asserted by Christensen et al. (2002; cited in Haneline & Young, 2009), it is recommended to consider this option as well.

Another possibly limiting factor in the study at hand is the fact that the global movements of the spine were not examined. In real-life practice a segmental examination is usually performed when the global examination shows an impairment, which could indicate a segmental dysfunction. A global examination was not performed in the study at hand, and it cannot be ruled out that there were only few or even no dysfunctions. This complicates the interpretation of the results. Arndt (2013) solved this problem in her study by using only participants with plated segments of the lumbar spine, guaranteeing that “a dysfunction” definitely existed.

The markings of the spinous processes, which were made to exclude the risk of wrong naming of the segment as a source of error, were partly blurred in the second round of examination and not clearly visible. In the study at hand the marks were made with a biro. The use of a water-resistant pen seems mandatory for future studies.

In the study at hand, the passive physiological motion palpation of the lumbar spine was carried out in sitting position. The sitting position was chosen, first, because this position is very often part of the examination in daily osteopathic practice, and second, also in view of the fact that previous studies using the sitting position were of poor quality. Compared to the lateral position, in the sitting position the surrounding tissue exerts an increased neuromuscular tension, which could make palpation more difficult. In daily osteopathic practice it is not enough to examine the lumbar spine in only one position; treatment decisions are usually made after examining the spine in different positions. The majority of previous reliability studies which performed examination in only one position showed poor interrater reliability (see above). The examination of the lumbar spine in different positions is more representative of daily osteopathic practice and should be considered for future studies, as already done by Qvistgaard et al. (2007).

The raters

Although from a statistical point of view no more than two raters are necessary (Patijn & FIMM Scientific Committee, 2004), this should be subjected to critical reflection. When comparing the findings of both raters in the study at hand (Table 5), it becomes obvious that rater 2 identified a significantly higher number of dysfunctions than rater 1. This could point to an outlier, which would mean that the results are not transferable to the entirety of osteopaths. Or it could be the result of the blurred markings of “normal” mobility and “hypomobility” (see above). Including more pairs of raters and groups of participants could improve the validity of similar studies.

The participants

22 participants is a rather low sample size which limits the validity of the study at hand. Furthermore, the participants were recruited via e-mail and SMS. It would have been more practical to work with patients consulting the osteopath on their own accord. Such participants would be more representative of daily osteopathic practice and would thus have enhanced the validity of the study at hand. For logistical reasons this was not possible. Since the raters and the assistants had their workplace in different locations, it was necessary to arrange one date for the examination of all subjects.

Prospects

Summing up the discussion above, following points should be considered for future studies:

- The implementation of a preliminary assessment (global movement) of the lumbar spine, with the aim to identify participants with/without dysfunctions in advance;

- The examination of the lumbar spine in different positions, to make it more representative for daily osteopathic practice;
- The necessity of the standardisation of the handhold, although it is not representative for the practice;
- The inclusion of more pairs of raters and groups of participants, to improve validity and identify possible outliers;
- The development of detailed standards for the interpretation of the findings, to minimize bias;
- The counting of two adjacent segments as agreement, as manipulation always affects a region and the relevance of the exact level of the dysfunction should be questioned;

CONCLUSION

The purpose of the study at hand was to test the interrater reliability of the specific segmental physiological motion palpation of the lumbar spine in sitting position. Before carrying out the study, the literature was searched for quality criteria of interrater reliability studies. Those were summarised and built the basis for the applied study design.

Two experienced raters examined 22 symptomatic and asymptomatic participants in sitting position. All directions of movement were tested and the most hypomobile segment was identified. The kappa values ranged from $k = 0.04$ to $k = 0.13$. These results confirm the results of earlier studies and lead to the conclusion that the interrater reliability of the specific segmental physiological motion palpation of the lumbar spine is poor.

The transferability of these result to the entirety of osteopaths must be questioned, since the results might be influenced by an outlier. Although it was attempted to standardise and objectify the examination and assessment procedures, in this kind of setting the findings will always be influenced by the subjective perception of the rater. However, a poor interrater reliability does not necessarily mean that this part of the examination is inappropriate. It only supports the critical reflection of the findings, which were gathered from only one test. In future studies, examination should be performed in different positions, as this is common in daily osteopathic practice.

The relevance of the exact level of dysfunction should be reflected on. After all, how can the efficacy of manipulation in the lumbar spine be explained when there is such poor reliability in identifying the dysfunction?

References

- Arndt, K. (2013). *Intertester-Reliabilität einer Kombination mehrerer manualtherapeutischer segmentaler Beweglichkeitstests: In Gegenüberstellung mit einem Test zur Identifikation einer Hypomobilität im Bereich der Lendenwirbelsäule*. Krems: Masterthese. Donau-Universität Krems.
- Binkley, J., Stratford, P. W., & Gill, C. (1995). Interrater Reliability of Lumbar Accessory Motion Mobility Testing. *Physical Therapy, 75*(9), 786–792.
- Bortz, J., & Lienert, G. A. (2003). *Kurzgefasste Statistik für die klinische Forschung: Leitfaden für die verteilungsfreie Analyse kleiner Stichproben* (2nd ed.). Heidelberg; Berlin u.a.: Springer.
- Christensen, H. W., Vach, W., Vach, K., Manniche, C., Haghfelt, T., Hartvigsen, L., & Høilund-Carlsen, P. F. (2002). Palpation of the upper thoracic spine: An observer reliability study. *Journal of Manipulative and Physiological Therapeutics, 25*(5), 285–292. doi:10.1067/mmt.2002.124424
- Degenhardt, B. F., Snider, K. T., Snider, E. J., & Johnson, J. C. (2005). Interobserver reliability of osteopathic palpatory diagnostic tests of the lumbar spine: improvements from consensus training. *JAOA: Journal of the American Osteopathic Association, 105*(10), 465–473.
- Deore, M., & May, S. (2012). The inter-rater and intra-rater reliability of passive physiological accessory movement assessment of lumbar spine in novice manual therapists. *Journal of Bodywork and Movement Therapies, 16*(3), 289–293. doi:10.1016/j.jbmt.2011.12.006
- Haneline, M., Cooperstein, R., Young, M., & Birkeland, K. (2008). Spinal motion palpation: A comparison of studies that assessed intersegmental end feel vs excursion. *Journal of*

Manipulative and Physiological Therapeutics, 31(8), 616–626.
doi:10.1016/j.jmpt.2008.09.007

Haneline, M., Cooperstein, R., Young, M., & Birkeland, K. (2009). An annotated bibliography of spinal motion palpation reliability studies. *The Journal of the Canadian Chiropractic Association*, 53(1), 40–58.

Haneline, M., & Young, M. (2009). A review of intraexaminer and interexaminer reliability of static spinal palpation: a literature synthesis. *Journal of Manipulative and Physiological Therapeutics*, 32(5), 379–386. doi:10.1016/j.jmpt.2009.04.010

Hestbaek, L., & Leboeuf-Yde, C. (2000). Are chiropractic tests for the lumbo-pelvic spine reliable and valid? A systematic critical literature review. *Journal of Manipulative and Physiological Therapeutics*, 23(4), 258–275.

Horneij, E., Hemborg, B., Johnsson, B., & Ekdahl, C. (2002). Clinical tests on impairment level related to low back pain: a study of test reliability. *Journal of Rehabilitation Medicine*, 34(4), 176–182.

Huijbregts, P. A. (2002). Spinal motion palpation: a review of reliability studies. *Journal of Manual & Manipulative Therapy*, 10(1), 24–39.

Johansson, F. (2006). Interexaminer reliability of lumbar segmental mobility tests. *Manual Therapy*, 11(4), 331–336. doi:10.1016/j.math.2005.06.014

Keating, J. C., Jr, Bergmann, T. F., Jacobs, G. E., Finer, B. A., & Larson, K. (1990). Interexaminer reliability of eight evaluative dimensions of lumbar segmental abnormality. *Journal of Manipulative and Physiological Therapeutics*, 13(8), 463–470.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174.

Maher, C., & Adams, R. (1994). Reliability of Pain and Stiffness Assessments in Clinical Manual Lumbar Spine Examination. *Physical Therapy*, 74(9), 801–809.

- May, S., Littlewood, C., & Bishop, A. (2006). Reliability of procedures used in the physical examination of non-specific low back pain: A systematic review. *Australian Journal of Physiotherapy*, (52), 91–102.
- Patijn, J., (Ed.), & International Federation for Manual/Musculoskeletal Medicine (FIMM) Scientific Committee. (2004). *Reproducibility and validity studies of diagnostic procedures in manual/musculoskeletal medicine*, Protocol formats, 3rd edition. Retrieved from http://www.fimm-online.com/pub/en/data/objects/reproducibility_validity.pdf
- Qvistgaard, E., Rasmussen, J., Laetgaard, J., Hecksher-Sorensen, S., & Bliddal, H. (2007). Intra-observer and inter-observer agreement of the manual examination of the lumbar spine in chronic low-back pain. *European Spine Journal*, 16(2), 277–282. doi:10.1007/s00586-006-0134-8
- R Core Team. (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing (Version 3.0.2). Vienna, Austria. Retrieved from <http://www.R-project.org/>
- Schneider, M., Erhard, R., Brach, J., Tellin, W., Imbarlina, F., & Delitto, A. (2008). Spinal palpation for lumbar segmental mobility and pain provocation: an interexaminer reliability study. *Journal of Manipulative and Physiological Therapeutics*, 31(6), 465–473. doi:10.1016/j.jmpt.2008.06.004
- Seffinger, M. A., Najm, W. I., Mishra, S. I., Adams, A., Dickerson, V. M., Murphy, L. S., & Reinsch, S. (2004). Reliability of spinal palpation for diagnosis of back and neck pain: a systematic review of the literature. *Spine*, 29(19), E413–E425.
- Shaw, K. A., Dougherty, J. J., Treffer, K. D., & Glaros, A. G. (2012). Establishing the content validity of palpatory examination for the assessment of the lumbar spine using ultrasonography: a pilot study. *JAOA: Journal of the American Osteopathic Association*, 112(12), 775–782.

- Sim, J., & Wright, C. C. (2005). The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Physical Therapy, 85*(3), 257–268.
- Snider, K. T., Snider, E. J., Degenhardt, B. F., Johnson, J. C., & Kribs, J. W. (2011). Palpatory accuracy of lumbar spinous processes using multiple bony landmarks. *Journal of Manipulative and Physiological Therapeutics, 34*(5), 306–313. doi:10.1016/j.jmpt.2011.04.006
- Stochkendahl, M. J., Christensen, H. W., Hartvigsen, J., Vach, W., Haas, M., Hestbaek, L., ... Bronfort, G. (2006). Manual examination of the spine: a systematic critical literature review of reproducibility. *Journal of Manipulative and Physiological Therapeutics, 29*(6), 475–485.e10. doi:10.1016/j.jmpt.2006.06.011
- Triano, J. J., Budgell, B., Bagnulo, A., Roffey, B., Bergmann, T., Cooperstein, R., ... Tepe, R. (2013). Review of methods used by chiropractors to determine the site for applying manipulation. *Chiropractic & Manual Therapies, 21*(1), 36. doi:10.1186/2045-709X-21-36
- Van Trijffel, E., Anderegg, Q., Bossuyt, P. M. M., & Lucas, C. (2005). Inter-examiner reliability of passive assessment of intervertebral motion in the cervical and lumbar spine: a systematic review. *Manual Therapy, 10*(4), 256–269. doi:10.1016/j.math.2005.04.008

Annex: raw data, 3rd may 2014

		Sagittal plane				Frontal plane				Transverse plane			
Participant	First rater	Flexion		Extension		Lateral flexion left		Lateral flexion right		Rotation left		Rotation right	
		Rater1	Rater2	Rater1	Rater2	Rater1	Rater2	Rater1	Rater2	Rater1	Rater2	Rater1	Rater2
1	2	0	0	5	0	0	0	1	0	0	0	0	0
2	2	0	2	0	5	2	5	5	5	2	5	1	0
3	2	3	4	3	3	2	4	3	1	3	2	3	0
4	2	1	1	2	2	3	3	2	2	2	1	2	1
5	2	0	1	0	1	0	1	1	1	0	1	3	1
6	2	0	5	0	5	2	5	0	5	0	5	2	5
7	2	0	5	0	0	0	5	3	1	0	5	0	1
8	2	0	2	0	2	0	0	0	2	0	2	0	2
9	2	0	0	0	0	0	0	0	0	2	1	2	1
10	2	0	2	0	1	0	1	0	1	4	1	4	2
11	1	0	1	3	0	0	0	0	1	2	1	2	1
12	2	4	4	4	4	0	5	0	1	4	3	4	4
13	1	4	1	4	1	0	4	0	5	0	4	2	5
14	2	0	4	0	1	0	4	0	4	0	1	1	1
15	1	0	1	0	1	5	4	4	4	5	3	4	2
16	1	0	4	4	0	3	4	0	5	0	4	0	5
17	1	0	3	3	1	4	3	4	3	0	3	3	3
18	1	0	5	0	1	0	3	3	2	2	2	3	2
19	1	3	3	3	2	0	4	4	3	4	4	0	4
20	1	4	5	0	1	0	3	0	1	0	1	2	1
21	1	0	1	2	1	0	1	4	0	2	4	2	1
22	1	0	5	0	1	4	0	0	4	0	4	4	4