

**Reliabilität der Testung
der abdominalen Diaphragmaspannung
im Seitenvergleich
(Interrater-Test und Test-Retest Studie)**

Master Thesis zur Erlangung des Grades
Master of Science in Osteopathie

an der **Donau Universität Krems –
Zentrum für chin. Medizin & Komplementärmedizin**

Wiener Schule für Osteopathie

von **Sonja Gruber**

Laakirchen, März 2013

Betreut von Dr. Erich Mayer-Fally

Eidesstattliche Erklärung

Hiermit versichere ich, die vorgelegte Masterthese selbständig verfasst zu haben.

Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten oder nicht veröffentlichten Arbeiten anderer übernommen wurden, wurden als solche gekennzeichnet. Sämtliche Quellen und Hilfsmittel, die ich für die Arbeit genützt habe, sind angegeben. Die Arbeit hat mit gleichem Inhalt weder im In- noch im Ausland noch keiner anderen Prüfungsbehörde vorgelegen.

Diese Arbeit stimmt mit der von dem/der Gutachter/in beurteilten Arbeit überein.

Datum

Unterschrift

Danksagung

Bei allen, die mich bei meiner Ausbildung zur Osteopathin und beim Erstellen der Masterthese unterstützt haben, möchte ich mich herzlich bedanken:

- Bei Dr. Erich Mayer-Fally, MSc D.O., für die schlussendliche Betreuung
- Bei Dr. Gebhard Woisetschläger für die guten Tips und die statistische Evaluation
- Bei meinem Bruder Robert Torlutter für die englische Übersetzung
- Bei meinem Mann für die vielfältige Unterstützung
- Bei meinen Eltern und meiner Schwiegermutter für die tolle Kinderbetreuung
- Bei meinen Kindern für die Zeit für meine "Schule"
- Bei Wolfgang Aspalter, MSc D.O., und Georg Heel als Osteopathen für meine herausfordernde Studie
- Bei den Probanden und meinen Arbeitskollegen als "Assistenten" bei der Studie
- Bei Andrea Greisberger MSc für die anregenden Gespräche und das engagierte Korrekturlesen des Konzepts und der Endfassung

Abstract (English Version)

Title: The Test of the Abdominal Diaphragm Tension and its Reliability concerning the Right and Left Body (Interrater-Test Retest Study)

Sonja Gruber, Vienna School of Osteopathy, 2013

Context: In Osteopathy the palpatory test of the diaphragm can be conducted in a static, respiratory, dynamic or inhibitoric way. The reliability of these taught and described tests has not been verified yet. As soon as a test turns out to be repeatable and independent from the tester it becomes a valuable diagnostic device for therapy concept and prognosis.

Objectives: The thesis surveys the interrater and test-retest reliability of the abdominal diaphragm tension in the left and right body. Furthermore the investigation shows whether the diaphragm test results correspond within patients with chronic pain in the lumbar, thoracolumbar and thoracic area.

Methods: The concept of the methodological study was chosen to answer the survey's questions. Two osteopaths examined totally 43 probands. 22 Probands showed no signs of pain and 21 test persons featured pain in the lumbar, thoracolumbar or thoracic area. All 43 probands took part in the interrater test. 20 of the 43 test persons were drawn for a second examination to check the test-retest reliability. A probationary examination was held before the main survey, the examiners were instructed and a randomization of the probands' order took place. The osteopath tested the diaphragm tension on the patient in a dorsal position by a translation movement in the frontal plane of the costal arch. The examiner judged whether the tension was stronger on any side or equal. Both the probands and the examiners were blinded.

Results: No interrater reliability ($\kappa=-0,35$) could be shown between the two examiners. However the Nonaccordance differs significantly from the expected coincident accordance ($p=0,00070$). Examiner 1 shows a fair accordance in the test-retest reliability ($\kappa=0,57$, $p=0,00095$), examiner 2 reveals no accordance significantly higher than coincidence ($\kappa=-0,02$, $p=0,91$). Furthermore no accordance higher than coincidence could be achieved between the judgements of the two testers and probands's self described side of pain ($\kappa=0,11$ for tester 1 and $\kappa=-0,16$ for tester 2). Both examiners revealed a preferred side for detecting tension (Tester 1 mainly found tensions in the left body, tester 2 in the right body).

Conclusion: The examiners reacted differently to tension in the two body sides. A higher number of examiners would probably allow generalizations about osteopaths as a whole.

Keywords: interrater reliability, test-retest reliability, abdominal diaphragm, tension, palpation

Abstract (Deutsche Version)

Titel: Reliabilität der Testung der abdominalen Diaphragmaspannung im Seitenvergleich (Interrater-Test Retest Studie)

Sonja Gruber, Wiener Schule für Osteopathie, 2013

Hintergrund: Die palpatorische Testung des Diaphragmas erfolgt in der Osteopathie statisch, respiratorisch, dynamisch oder inhibitorisch. Keiner der unterrichteten und beschriebenen Tests wurde bislang auf seine Zuverlässigkeit geprüft. Erfüllt ein Test seine Wiederholbarkeit und Unabhängigkeit vom Tester, ist er ein brauchbarer Baustein zur Diagnose zum Therapiekonzept sowie zur Prognose.

Ziele: Ziel ist die Prüfung der Interrater- und Test-Retest Reliabilität der abdominalen Diaphragmaspannung im Seitenvergleich. Weiters wird untersucht, ob die Ergebnisse des Diaphragmatests bei Patienten mit chronischen Schmerzen im lumbalen, thorakolumbalen oder thorakalen Bereich statistisch signifikant übereinstimmen.

Methode: Zur Beantwortung der Studienfragen wurde das Design der methodologischen Studie gewählt. Insgesamt 43 Probanden, davon 22 Probanden ohne Schmerzsymptomatik und 21 Schmerzpatienten (lumbal, thorakolumbal oder thorakal), wurden von zwei Osteopathen untersucht. Alle 43 Probanden nahmen an der Interrater Testung teil, 20 wurden von den 43 für die Test-Retest Reliabilität zur zweiten Untersuchung ausgelost. Vor der Hauptuntersuchung fand eine Probedurchführung, eine Einschulung der Untersucher und eine Randomisierung der Probandenreihenfolge statt. Die Diaphragmaspannung wurde in Rückenlage mit einer vom Osteopathen induzierten Translationsbewegung in der Frontalebene am Rippenbogen in Ruheatmung getestet. Der Untersucher beurteilte, ob die Spannung beidseits gleich, rechts oder links unterschiedlich ist. Sowohl die Probanden als auch die Untersucher waren blindiert.

Ergebnisse: Zwischen den beiden Untersuchern besteht keine Interrater Reliabilität ($\kappa = -0,35$), wobei deren Nicht-Übereinstimmung signifikant von einer durch Zufall zu erwartenden Übereinstimmung abweicht ($p = 0,00070$). Die Test-Retest Reliabilität ist bei Tester 1 mittelmäßig ($\kappa = 0,57$, $p = 0,00095$), bei Tester 2 ist keine Übereinstimmung, die über das Ausmaß einer zufälligen Übereinstimmung hinausgeht zu beobachten ($\kappa = -0,02$, $p = 0,91$). Zusätzlich geht bei beiden Untersuchern die Übereinstimmung der Untersuchungsergebnisse mit der aufgrund der angegebenen Schmerzseite des Probanden erwarteten Körperseite nicht über den Zufall hinaus (Tester 1 mit $\kappa = 0,11$, Tester 2 mit $\kappa = -0,16$). Als signifikant stellte sich bei beiden Untersuchern eine bevorzugte Spannungsseite heraus (Tester 1 fand vorwiegend Spannungen in der linken, Tester 2 in der rechten Körperhälfte).

Konklusion: Die Untersucher reagieren unterschiedlich auf die Spannungen jeder Körperseite der Probanden. Eine größere Anzahl an Untersuchern würde ein Verallgemeinern auf die Gesamtheit der Osteopathen eher zulassen.

Schlüsselwörter: Interrater Reliabilität, Test-Retest Reliabilität, abdominales Diaphragma, Spannung, Palpation

Inhaltsverzeichnis

Eidesstattliche Erklärung	II
Danksagung	III
Abstract (English Version)	IV
Abstract (Deutsche Version)	V
Inhaltsverzeichnis	VII
1. Einleitung	1
1.1 Literaturrecherche	2
2. Theorieteil	3
2.1 Palpation als Basis für die osteopathische Diagnosestellung.....	3
2.2 Reliabilität in der osteopathischen Wissenschaft	5
2.3 Das abdominale Diaphragma	7
2.3.1 Das Diaphragma, seine Aufgabe und Integration im Körper	9
2.3.2 Dysfunktion und Untersuchung vom Diaphragma.....	12
3. Empirischer Teil	16
3.1 Definition der Testung der abdominalen Diaphragmaspannung.....	16
3.2 Forschungsfragen und Hypothesen	17
3.2.1 Hypothese 1	17
3.2.2 Hypothese 2.....	17
3.2.3 Hypothese 3.....	17
3.3 Forschungsdesign	18
3.4 Stichprobenbeschreibung	19
3.4.1 Untersucher (Osteopathen)	19
3.4.2 Probanden	19
3.4.3 Einschlusskriterien	20
3.4.4 Ausschlusskriterien	20
3.4.5 Demographische Angaben.....	20
3.5 Studienablauf	21
3.5.1 Probedurchlauf.....	21
3.5.2 Einschulung der Osteopathen	23
3.5.3 Randomisierung.....	24

3.5.4	Untersuchungsablauf	25
3.6	Statistische Datenverarbeitung und -auswertung.....	27
3.6.1	Grundlegendes	27
3.6.2	Durchführung der statistischen Auswertung	27
4.	Ergebnisse	30
4.1	Interrater Reliabilität	30
4.2	Test-Retest Reliabilität	32
4.3	Schmerzangaben und Untersuchungsergebnisse im Vergleich	33
4.4	Ursachenforschung für die systematischen Unterschiede.....	36
4.4.1	Ergebnisse der Feedbackbögen der Probanden	36
4.4.2	Statistische Untersuchung eines möglichen Behandlungseffekts	37
4.4.3	Statistische Untersuchung von Einflüssen der Räumlichkeiten.....	39
4.4.4	Häufigkeit der Ergebnisse in Abhängigkeit vom Tester	40
4.5	Überprüfung der Hypothesen	41
4.5.1	Hypothese 1	41
4.5.2	Hypothese 2.....	41
4.5.3	Hypothese 3.....	42
5.	Diskussion	43
5.1	Diskussion zur Hypothese 1	43
5.2	Diskussion zur Hypothese 2	45
5.3	Diskussion zur Hypothese 3	46
5.4	Kritik an der eigenen Studie.....	47
5.5	Ausblick.....	48
6.	Schlussfolgerung	50
7.	Abkürzungsverzeichnis	52
8.	Abbildungsverzeichnis	53
9.	Tabellenverzeichnis	54
10.	Literaturverzeichnis	55
11.	Anhang	61
	<i>Ärzteinformation</i>	<i>61</i>
	<i>Probandeninformation</i>	<i>62</i>
	<i>Fragebogen für die Probanden.....</i>	<i>63</i>
	<i>Feedbackbogen Probanden</i>	<i>64</i>

<i>Probandentabelle</i>	65
<i>Originale Eintragungen der Untersuchungsergebnisse</i>	66
<i>Reihenfolge gesamt</i>	70
<i>Reihenfolge Proband</i>	71
<i>Interreliabilität im Testdurchgang 1</i>	72
<i>Test- Retest Reliabilität im Testdurchgang 2</i>	73
<i>Untersuchungsreihenfolge von Tester 1 und 2</i>	74
<i>Mögliche Einflüsse der unterschiedlichen Untersuchungsräume</i>	76
<i>Häufigkeit der Untersuchungsergebnisse in Abhängigkeit der Tester</i>	77
12. Summary	78

1. Einleitung

In meiner Ausbildungszeit als Osteopathin haben mich die unterschiedlichen Herangehensweisen von den verschiedenen Vortragenden immer wieder fasziniert. Zum anderen überforderten mich die vielen gleichzeitig möglichen Wahrheiten für das Ziel Gesundheit.

Mit dem Kennenlernen der verschiedenartigen Erscheinungsbilder einer definierten Krankheit wurde mir der Vorteil der Osteopathie durch die Orientierung an der Individualität des Patienten bewusst. Diese Individualität des Patienten zeigt sich in den jeweiligen Gewohnheiten, Strukturen und weiterlaufenden Mechanismen im gesamten Körper. Das Eingehen auf die verschiedenen Erscheinungsbilder in Bezug zur Individualität des Patienten macht den Reiz der Suche nach Gesundheit im Rahmen einer osteopathischen Behandlung aus.

Der Individualität des Patienten wird von vielen Osteopathen mit einer persönlichen Freiheit der Untersuchung und Behandlung begegnet. Die Individualität des Osteopathen ist ein dehnbarer Begriff und maßgeblich von der Intuition desselben beeinflusst. Albert Einsteins Aussage, "Phantasie ist wichtiger als Wissen, denn Wissen ist begrenzt" (Viereck, 1929), drückt für mich die Wichtigkeit der Intuition für den Erfolg in der Behandlung aus.

Trotz allem findet sich in mir neben der Wertschätzung für diese persönliche Freiheit auch der Wunsch nach Standardisierung der Untersuchung und des Behandlungsansatzes, um diese vergleichbar zu machen. Die Vergleichbarkeit einer Untersuchung innerhalb eines Testers und zwischen verschiedenen Testern kann als Wiederbefundparameter, als Kommunikationsmittel und als verlässliche Basis für die Behandlung herhalten.

Die Auswahl der Untersuchungen und die Wahl auf einen bestimmten Behandlungsansatz hängt von vielen Facetten (Ausbildungsstand des Osteopathen, Prioritätensetzung der Beschwerden vom Patienten, Interaktion Osteopath-Patient, Persönlichkeit vom Osteopathen, vom Patienten...) ab.

In meiner Masterarbeit reduziere ich mich auf einen einzelnen diagnostischen Test um eine fundierte Prüfung der Interrater Reliabilität und der Test-Retest Reliabilität durchführen zu können. Ein Test soll einen Fokus für die Behandlung schaffen und das Auswählen von effektiven Techniken erleichtern (Fritz & Wainner, 2001).

Tests zur abdominalen Diaphragmaspannung werden angesichts seiner Aussagekraft zum gesamten körperlichen Zustand des Patienten in der osteopathischen Anfangsuntersuchung häufig angewendet (Pope, 2003; Helsmoortel, 2002, S.91; Meert, 2007, S.323). Das abdominale Diaphragma hat aufgrund seiner federführenden Funktion in der Atmung, aber auch durch die zentrale Lage und der Pumpfunktion für die Flüssigkeiten einen Einfluss auf

den gesamten Körper (Meert, 2007, S.323; Hartmann, 2005, (2)S.54; (3)S.80; Helsmoortel, 2002, S.73) und wurde daher als Ort dieser methodologischen Studie gewählt.

In der vorliegenden Masterthese wird zuerst die Basis der osteopathischen Untersuchung, die Palpation, besprochen, um sie dann im Kontext der osteopathischen Forschung zu betrachten. Nach der theoretischen Auseinandersetzung mit dem Diaphragma und der Testung der abdominalen Diaphragmaspannung erfolgt eine Überleitung zum empirischen Teil. Der für diese Studie ausgewählte dynamische Diaphragmatest wird auf seine Interrater- und Test-Retest Reliabilität überprüft und kritisch hinterfragt.

Zur Erleichterung der Lesbarkeit wählte ich die maskuline Form für Untersucher, Probanden und Patienten. Selbstverständlich sind jeweils Personen aller Geschlechter gemeint.

1.1 Literaturrecherche

Für die Ausarbeitung der Studie wurden von März 2012 bis März 2013 folgende Begriffe, eingeschränkt auf Veröffentlichung nach dem 01.01.2000, in den Datenbanken Medline, JOAO, Google Scholar, Pubmed, Osteopathic Research und Science Direkt, eingegeben:

- Diaphragme (function, human, examination, abdominal, transversal)
- osteopathy
- reliability (interexaminer, intraexaminer)
- torsion (Common Compensatory Pattern), translation

Ebenso wurden geeignete Fachbücher mit Stand 2011 besorgt und folgende Experten befragt:

- Rittler Margit, MSc D.O., die ebenso eine methodologische Studie durchführte.
- Maier Olivia, MSc D.O., und Stemeseder Harald, MSc D.O., die an der WSO "Diaphragma und Lunge" unterrichten.
- Greisberger Andrea, MSc, die an der FH Wien Physiotherapie unterrichtet.
- Emaildialoge mit Prof. Paul Klein, der selbst schon viele Studien durchgeführt hat.
- Literaturempfehlungen von unterrichtenden Osteopathen an der WSO (Georges Finet, Christian Fossum, Dr. Erich Mayer-Fally, Jon Parsons, Christian Williame).

2. Theorieteil

Die Palpation als Grundlage der osteopathischen Diagnose stellt den ersten Teil des Theorieteils dar. Die Überleitung zur wissenschaftlichen Überprüfung einer Untersuchung führt zum Aufbau der Forschung im osteopathischen Feld, im besonderen zur Reliabilitätstestung. Der Untersuchungsgegenstand "abdominales Diaphragma" wird in seiner zentralen Funktion als Bindeglied zum gesamten Körper im darauffolgenden Teil beschrieben. Über die Dysfunktionen schließt sich dann der theoretische Bogen mit der Vorstellung der Tests der abdominalen Diaphragmaspannung.

2.1 Palpation als Basis für die osteopathische Diagnosestellung

Die Diagnose (griech. *Diagnosis* = Unterscheidung, Entscheidung) entsteht in der Medizin durch Aufnehmen, Erkennen, Vergleichen und Interpretieren der Information von und über den Patienten (Croibier, 2006, S.1-8; Angele, 2011). Weiters ist die Diagnose Ausgangspunkt für Therapiekonzept, Therapie und Prognose (Mayer-Fally, 2010, S.62). Jeder Teilbereich in der Medizin gewichtet und interpretiert die Symptome in ihrer eigenen Art und Weise, womit sich der dazu passende und je nach Ausbildungsstand mögliche Behandlungsansatz ergibt. Fritz & Wainner (2001) beschreiben die Diagnose als "the link between examination, findings and interventions".

Die osteopathische Diagnose wird von den Informationen aus der Anamnese, der Inspektion und der manuellen Untersuchung geformt. Die manuelle Untersuchung ist ein abgestimmtes Konglomerat aus Tests der Überprüfung der Position, der Beweglichkeit, der Kraft und Schmerzprovokation, aber auch der Palpation der Gewebequalität (Tonus, Gewebswiderstand, Bewegungstendenz). Glover et al (2006) beschreiben die Palpation als Auflegen der Finger auf die Hautoberfläche, wo mit variierendem Druck die darunterliegenden Gewebeszustände erfasst werden. Die Palpationstiefe wird bestimmt durch die Druckstärke, Griffhaltung und Intention (Liem, 2006, S.151).

Die Frage, was nun den Unterschied zwischen gesundem Gewebe und einer Dysfunktion ausmacht, führt in die Zellphysiologie und Biomechanik.

Der Begriff Dysfunktion wird für die verschlechterte oder veränderte Mobilität von skelettalen, myofascialen oder visceralen Strukturen mit deren vaskulären, lymphatischen und neurologischen Elementen verwendet (Glover et al, 2006).

Ausgehend von einem Entzündungsgeschehen oder einer Ischämie kommt es durch die Ausschüttung von Mediatoren wie Arachnidonsäure, Adenosintriphosphat, Oxidradikale usw. zu einer Reizung der Nozizeptoren (Mayer-Fally, 2010, S.28). Im normalen Heilungsgeschehen kommt es mit der folgenden gesteigerten lokalen Durchblutung zu einer

Ausschwemmung der Endprodukte der Entzündungstoffe und in der anschließenden Proliferationsphase zur Neustrukturierung der Bindegewebszellen, Fibroblasten und dergleichen mehr (Van den Berg et al, 2001, S.137). Übersäuerung des Gewebes, Medikamente, Stress, Rauchen, Wärmeanwendungen oder fehlende Belastungsreize können den Heilungsprozess stören und die Bildung von pathologischen Crosslinks, Kapselverklebungen oder Muskelverkürzungen fördern (Van den Berg & Cabri, 1999, S.280). Abb.1 zeigt die Einschränkungen bei Belastung im Bindegewebe im Normalzustand im Vergleich mit pathologischen Crosslinks.

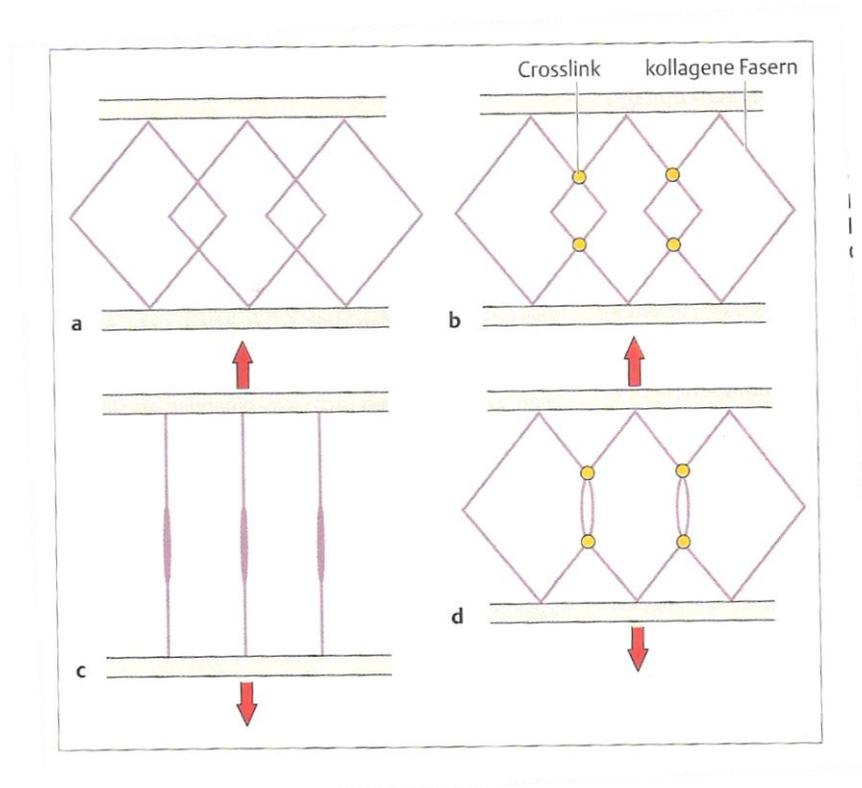


Abb. 1: Schematische Darstellung eines kollagenen Netzwerks (a) im entspannten Normalzustand, (b) entspannt mit pathologischen Crosslinks, (c) im belasteten Normalzustand und (d) belastet mit pathologischen Crosslinks (Van den Berg & Cabri, 1999, S.179).

Palpatorisch sind diese Gewebsveränderungen mit einer Asymmetrie, Tension und einer qualitativen wie quantitativen Mobilitätsveränderung spürbar, ebenso wird eine Sensibilitätsveränderung beschrieben (Glover et al, 2006; Mayer-Fally, 2010, S.32; Kuchera, 2007). Snider, Johnson & Degenhardt (2008) bestätigten diese Aussagen in ihrer Studie mit 16 chronischen "Low back pain"- Patienten und 47 Probanden ohne Schmerzen. Gewebsunterschiede waren bei den chronischen Patienten im Lumbalbereich im Vergleich zu den schmerzfreen Probanden signifikant häufiger vorhanden.

Biomechanisch betrachtet entsteht unter Einwirkung einer Kraft auf einen Körper entweder eine Bewegung des gesamten Körpers oder eine Bewegung innerhalb des Körpers. Die entstandene Deformierung des Körpergewebes bei Bewegung innerhalb des Körpers ist einerseits abhängig von der Dauer, der Größe, der Richtung und der Geschwindigkeit der

einwirkenden Belastung, andererseits von den verschiedenen chemischen, mechanischen und biophysischen Gewebeeigenschaften (Klein & Sommerfeld, 2004, S.104; Meert, 2007, S.80). Eine Studie von Feipel, Rondelet, Le Pallec, De Witte & Rooze (1999) zeigt den Unterschied von Patienten im Nackenbereich und asymptomatischen Probanden in der Harmonie des Bewegungsablaufs qualitativ und quantitativ.

Die Dysfunktion kann also nicht an einem Parameter festgemacht werden. Für eine probate Diagnose ist eine Zusammenschau mehrerer Tests vonnöten, um die Dysfunktion in ihrer breiten Facette greifbar zu machen. "Nur durch die Konvergenz der klinischen Analyse und der Kohärenz der Resultate mehrerer Tests kann eine zuverlässige Diagnose erstellt werden" (Croibier, 2006, S.57).

Statische und dynamische Tests sind die großen Kategorien der osteopathischen Tests. Bei den statischen Tests, auch *Écoute* oder Listening genannt, wird die Position und der Tonus beurteilt. Die dynamischen Tests nutzen die vom Untersucher induzierte Bewegung. Eine detailliertere Beschreibung der osteopathischen Tests in Bezug zum Diaphragma ist im Kapitel 2.3.2. nachzulesen.

Aussagekräftig für Praxis und Forschung sind palpatorische Tests, wenn die Ergebnisse unabhängig vom Tester wiederholbar sind. Eine Interrater- und Test-Retest Reliabilitätsstudie setzt genau bei dieser Fragestellung an.

2.2 Reliabilität in der osteopathischen Wissenschaft

Die Hierarchie der allgemeinen Forschung ergab sich aus dem historischen Kontext mit den Philosophen Sokrates und Platon, später deren ideellen Nachfolgern Newton, Decartes und Kant, die den Verstand ehrten und den Sinnesempfindungen misstrauten (Sidler, 2012).

Das mathematisch Beweisbare und Reproduzierbare setzte sich in der Forschung auch in der Medizin durch. So entstand eine Hierarchie gekrönt von Metaanalysen und systemischen Reviews, gefolgt von systematischen Reviews und blindierten Placebo kontrolliert randomisierten Studien, Designs mit Messwiederholungen, methodologische Studien und schlussendlich Einzelfallstudien (Walach, Falkenberg, Fonnebo, Lewith & Jonas, 2006).

Die Evidence Based Medicine, die diese Hierarchie achtet, kann am ehesten für begrenzte Fragen der Wirksamkeit von Medikamenten eingehalten werden (Walach et al, 2006).

Jaspers Aussage, das Unheil menschlicher Existenz beginne, wenn das wissenschaftlich Gewusste für das Sein selbst gehalten wird, und alles nicht wissenschaftlich Zugängliche gelte als nicht existent (Wisser, 1995, S.73), trifft das Regiment von wissenschaftlich Untersuchtem im Vergleich zum Erfahrungswissen.

Walach et al (2006) erwähnt neben der Osteopathie auch die Bereiche Physiotherapie, Akupunktur, Homöopathie und Chirurgie, für die die Wertung der Methode im hierarchischen Kontext nicht angebracht ist. Vor allem in Erfahrungswissenschaften, zu denen die Autorin

die Osteopathie zählt, ergeben sich spezifische Fragen aus der Praxis, die nicht mit verallgemeinernden Reviews, sondern mit spezifisch abgestimmten Methoden beantwortet werden sollen.

Um den Forschungsrichtlinien gerecht zu werden und trotzdem Aussagekraft für die Osteopathie zu erhalten, dürfen nach Walach et al (2006) vorrangig Patientenzufriedenheit, Individualität des Untersuchers und des Patienten Thema medizinischer Forschung sein. Sicherheit, Wirksamkeit und Kosten sind dabei als sekundär anzusehen.

Walach et al (2006) schlägt als Maßstab ein zirkuläres Modell im Sinne der Evidenced Practised Medicine vor. Wie in Abb.2 ersichtlich gibt es für die Evidenz vier verschiedene Kategorien (Effizienz, Effektivität, vergleichende Effektivität und Sicherheit). Im äußeren Kreis der Abbildung werden die dazu passenden Untersuchungsmethoden angeführt.

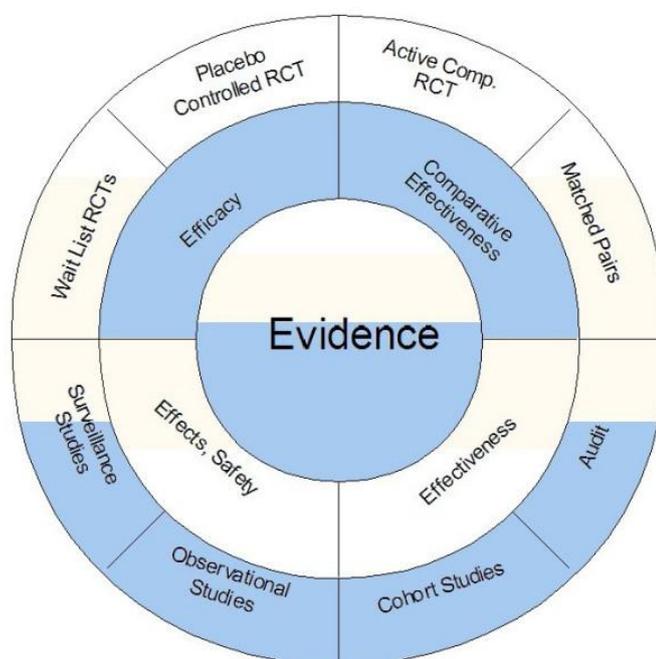


Abb.2: Zirkuläres Modell für evidence practised medicine (Walach et al, 2006)

Liem, Hilbrecht & Schmidt (2012) ergänzen die Aussage von Walach, indem sie nicht Wissensgewinn und Systematisierung von Wissen als vorrangige Forschungsfragen sehen, die in der klinischen Praxis weiterhelfen. Aus den Studien von Liem et al (2012), Johnston (1982), und Walach et al (2006) lässt sich folgendes nützliches Forschungsspektrum für die Osteopathie zusammenfassen:

- Verantwortliches Handeln am Einzelnen (Einzelfallstudien)
- Respektieren der Zweierkonstellation "Untersucher-Patient" (Qualitative Analysen, Designs mit Messwiederholungen)
- Wirkungsweisen osteopathischer Behandlungen und Effekte bestimmter Behandlungsstrategien (waiting list design)
- Überprüfen der Untersuchungsmethoden (Reliabilitätsstudien)

Nach Forschung über die Effizienz einer osteopathischen Behandlung bei verschiedenen Krankheitsbildern schlägt Liem et al (2012) also Intertester Reliabilitäten in der Befunderhebung als Gegenstand der Untersuchung vor.

Die Herausforderung für Reliabilitätsstudien liegt bei mehreren Parametern:

- Interaktion und Individualität von Untersucher und Proband
- Behandlungseffekt allein durch die Untersuchung
- Finden der gleichen Palpationstiefe und des gleichen Palpationsorts am Probanden.

Auf Grund der beschriebenen Problematik von methodologischen Untersuchungen setzt eine Interrater und Test-Retest Reliabilitätsstudie den Fokus auf eine überlegte und standardisierte Testausführung. Eine Reliabilitätsstudie gibt Hinweise auf die Zuverlässigkeit eines Tests, womit die Diagnose in seiner Allgemeinheit nachvollziehbarer werden kann. Sie kann mit weiteren Validitäts- und experimentell kontrollierten Studien zu einer umsichtigen Auswahl aus der "Testbatterie" helfen, so den Kern des Beschwerdebilds der Patienten in der Diagnose besser treffen und einen effektiveren Behandlungsansatz bieten.

2.3 Das abdominale Diaphragma

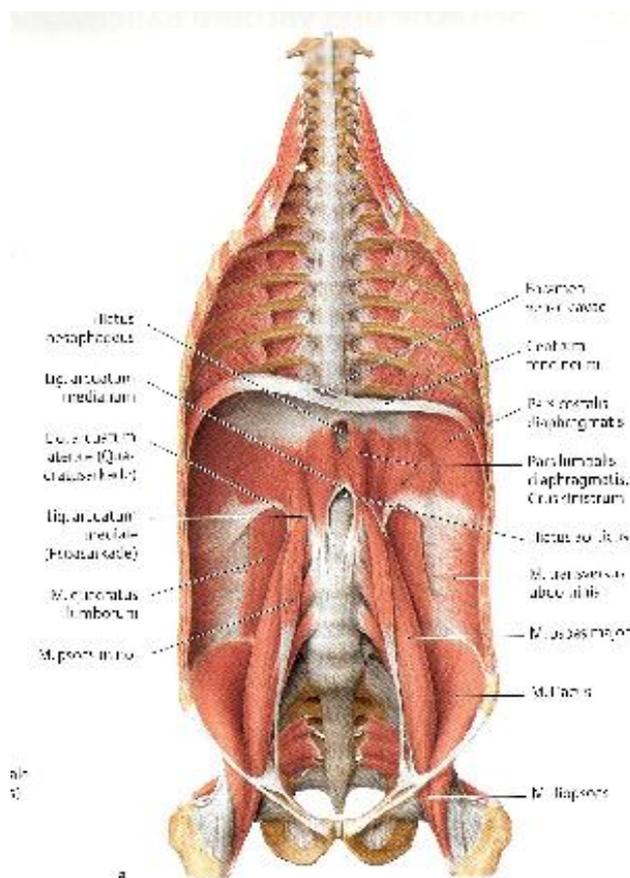


Abb.3: Abdominales Diaphragma (Schünke, 2005, S.147).

Wie in Abb. 3 ersichtlich trennt das abdominale Diaphragma mit einer Dicke von ca. 3mm den Brust- vom Bauchraum. Das Zwerchfell breitet sich vom Centrum Tendineum mit einem flächigen Ansatz wie folgt aus (Schünke, Schulte, Schumacher, Voll & Wesker, 2005, S.152):

- an den oberen Lendenwirbeln (Pars Lumbalis)
- an den unteren sechs Rippen (Pars Costalis)
- am Proc. Xiphoideus des Sternums (Pars Sternalis)

Etliche "Körperleitungen" finden durch die aufgespannte Kuppel in der Körpermitte einen Durchlass, wie es in Abb.3 und Abb.4 zu sehen ist: Durch das Foramen Venae Cavae in Höhe T8 beim Centrum tendineum treten die V. Cava Inferior und der N. Phrenicus. Beim Hiatus Oesophagus in Höhe T10 finden der Oesophagus und der N. Vagus den Durchgang. Der Hiatus Aorticus in Höhe T12 ist direkt vor dem Wirbelkörper die Öffnung für die Aorta und den Ductus Thoracicus. In Spalten vom Crus Mediale (Psoasarkade) durchdringen die V. Azygos, V. Hemiazygos und Nn. Splanchnici das Diaphragma, während der Truncus Sympaticus zwischen Spalten vom Crus Mediale und Crus Laterale (Quadratusarkade) den Weg findet. Das Trigonum Sternocostale beim Pars Sternalis bietet Durchgang für die A. und V. Thoracica Interna (Leonhardt, Tillmann, Töndury & Zilles, 1987, S.304; Schünke et al, 2005, S.90).

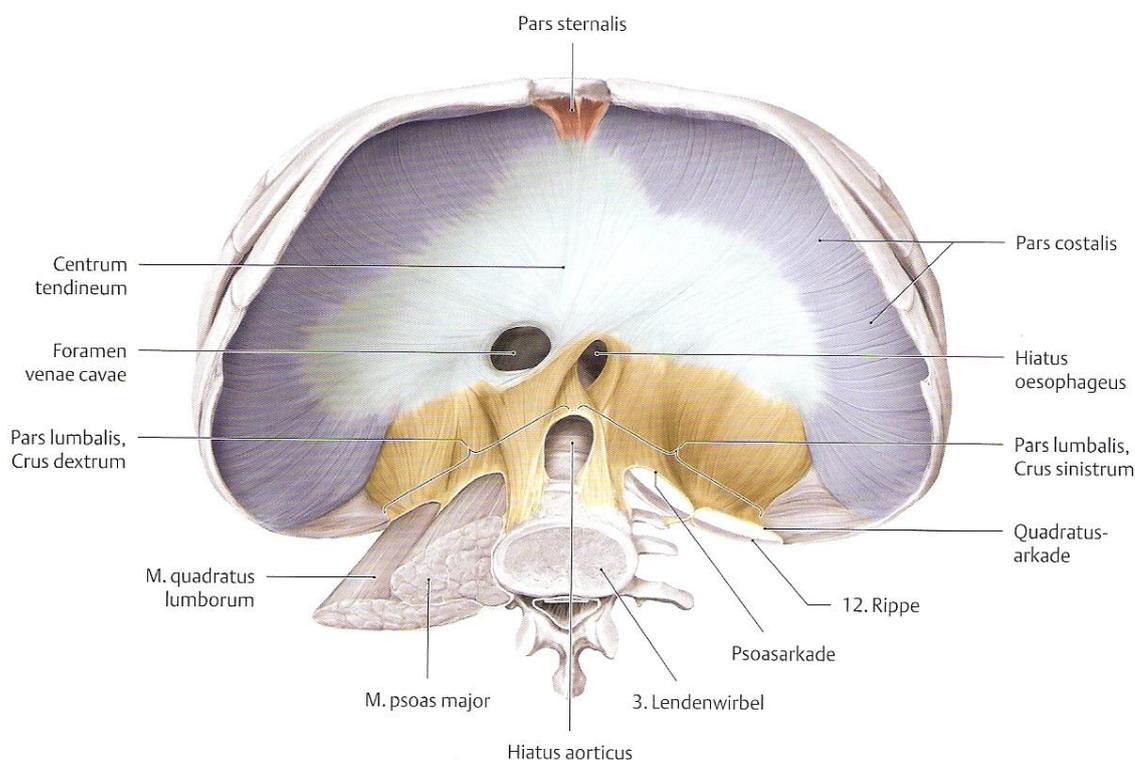


Abb.4: Diaphragma im transversalen Überblick (Schünke, 2005, S. 134).

Von verschiedenen Autoren (Liem, 2005, S.457; Greenman, 2005, S.10; Meert, 2007, S.140) werden ligamentöse Verbindungen beschrieben:

- zum Herz (Lig. Phrenicopericardica)
- zur Lunge (Lig. Phrenicopleurale)
- zur Leber (Lig. Triangulare, Lig. Coronarium, Lig. Teres Hepatis)
- zum Duodenum (M. von Treitz)
- zum Magen (Lig. Gastrophrenicum)
- zum Dickdarm (Lig. Phrenicocolicum)
- zur Niere (Fascia Renalis, Fascia Retrorenalis)

2.3.1 Das Diaphragma, seine Aufgabe und Integration im Körper

Die Kontraktion bei der Inspiration bewirkt, wie in Abb.5 ersichtlich, eine Abflachung und Senkung der Zwerchfellkuppel. Der Unterdruck des Thorax wird größer, die Bauchorgane werden vom Diaphragma angesaugt. Der positive abdominale Druck vergrößert sich und der Bauch wird nach unten und vorne gedrückt. Der Beckenboden bietet dabei die Gegenspannung.

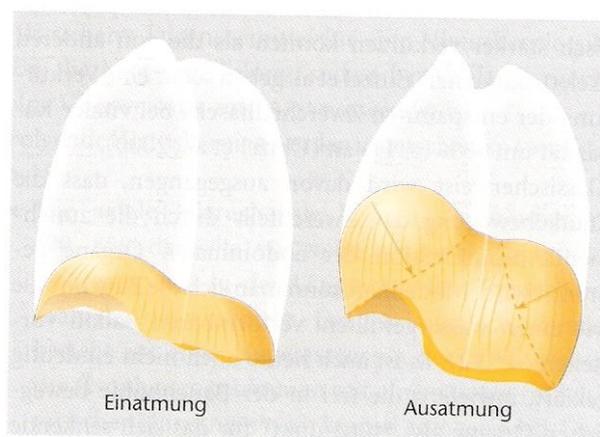


Abb.5: 3D-Rekonstruktion des Diaphragmas, Ansicht von ventral in maximaler Einatmung bzw. Ausatmung (Meert, 2007, S.136).

Durch den Wechsel der Kräfte in der Ein- und Ausatmung entsteht ein rhythmischer Pumpeffekt, von dem viele Körpergewebe profitieren - so beschrieben in anatomischen und osteopathischen Lehrbüchern (Leonhardt et al, 1987, S.306; Helsmoortel, 2002, S.79; Brazzo 2004, S.20, Meert, 2007, S.135). Als "zweites venöses Herz" (Brazzo, 2004, S.20) bewirkt das Diaphragma eine Sogwirkung auf venöse, lymphatische und interstitielle Flüssigkeiten. Außerdem wird durch die Tätigkeit des Diaphragmas die Peristaltik erleichtert. Bei Husten, Lautbildung, Defäkation und Entbindung spielt das Diaphragma eine basale Rolle (Liem, 2005, S.458, Meert, 2007, S.123).

Untersuchungen zum Diaphragma und seine Aufgaben ergaben, neben der Hauptleistung Inspiration, weiter interessante Details: Im Review von Pickering & Jones (2002) konnte beim Schlucken und Erbrechen im Tierversuch mit Ratten dem costalen Teil eine abdominale Druckerhöhung und dem cruralen Anteil eine Weiterleitung des Drucks an den

Ösophagusphinkter zugeteilt werden. Diese Druckweiterleitung ist insofern bemerkenswert, da sie ein Zusammenspiel von Inspirations-, Halte- und Peristaltikmuskulararbeit zeigt.

Hodges & Gandevia (2000) zeigten in ihrer Studie an vier Menschen mittels wiederholender Armbewegung die vermehrte EMG-Aktivität des Diaphragmas bei Haltearbeit unabhängig von der Phase der Atmung im Sitzen und im Stehen. So wird der intraabdominale Druck durch die Kontraktion des Diaphragmas erhöht, damit könnten die Bauchmuskeln die spinale Stabilität über eine Spannung der Fascia thorakolumbalis erhöhen.

Whitelaw (1987) dokumentierte anhand von CT-Aufnahmen eines Menschen, dass die typische Sattelform des Zwerchfells den Hemidiaphragmen ermöglicht, unabhängig voneinander zu arbeiten. Die rechte Zwerchfellkuppel verstellte sich craniocaudal um 6,7-7,2 cm und die linke um nur 4,0-4,3 cm. Ünal, Arslan, Uzun, Özbay & Sakaraya (2000) bestätigten die unabhängige Ausdehnung der Hemidiaphragmen bei gesunden Probanden (rechts 69mm, links 56mm) und bei chronisch obstruktiv pulmonalen Patienten (rechts 26mm, links 20mm). Ein seitendifferenzierter Tonus der Hemidiaphragmen wäre also durchaus plausibel.

In osteopathischen Fachbüchern (Liem, 2005, S.458; Meert, 2007, S.123) wird das Modell der Muskelketten beschrieben: ventrale und dorsale Muskelketten kreuzen sich in der Körpermitte, also im Diaphragma. Dadurch entsteht für das Diaphragma eine Schlüsselposition bei der Bewältigung von dynamischen und statischen Aktivitäten.

Nach Myers (2004, S.108) und Paoletti (2001, S.60) läuft die Kraftübertragung vom M. Pectoralis Major, gekreuzt in der Aponeurose, der Faszia Transversalis und im Diaphragma zum kontralateralen M. Obliquus Internus und in der Fortsetzung nach dorsal zum M. Gluteus Medius (siehe auch Abb.6).

Die Muskelschlinge vom M. Rhomboideus dorsal zum M. Serratus Anterior findet seine Fortsetzung ventral im M. Obliquus externus und über das Os Pubis zu den Adduktoren kontralateral. Hypertonus am Rippenbogen oder am Diaphragma kann sich demnach nach cranial, caudal und/oder kontralateral fortsetzen.

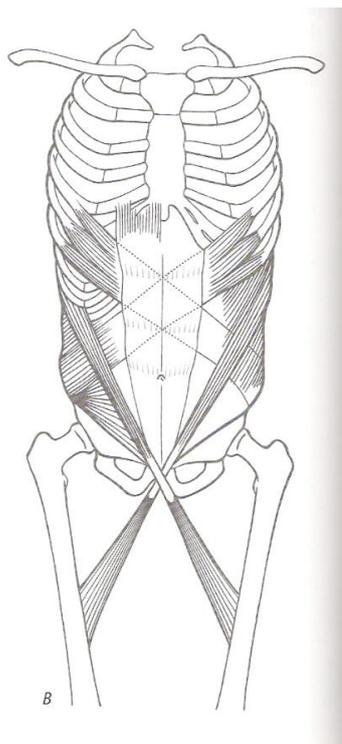


Abb.6: Kreuzung der Körperseite durch M. Obliquus Externus und Adduktoren, M. Obliquus Internus und M. Gluteus Medius (Myers, 2004, S.180).

Die Thorakolumbalen Faszie (TFL) überzieht den M. Erector Spinae und formt außerdem die Aponeurose der Bauchwand. Dies kann laut Benjamin (2009) der Grund sein, warum Kräfte, die die Spannung in der TFL verändern, auf die BWS und die Rippen weitergegeben werden können. Eine weitere Verbindung der TFL besteht zur Glutealmuskulatur, womit die Spannungsweitergabe an das Becken und zur unteren Extremität, auch zur kontralateralen Seite, erklärt werden könnten. Die beschriebenen Spannungsweiterleitungen wurden bei den Aktivitäten Laufen und Schwimmen gemessen.

In der osteopathischen Literatur stößt man beim Stichwort Integration immer wieder auf das Tensegritymodell. "Tensegrity" beschreibt die mögliche Weiterleitung und Verbindung der Kräfte (Kompression und Dekompression) auf Muskeln, Sehnen, Ligamente und Knochen, da der gesamte Körper mit Flüssigkeit, Faszien und neuronaler Aktivität durch Spannung in Verbindung zueinander steht (Chen & Ingber, 2008; Pflüger, 2010). Dieses Modell wurde einstweilen nur im Unterarm vom Chirurgen Guimberteau et al (2008) histologisch bestätigt. Die folgenden Untersuchungen von Kary und Finet sprechen von der Weiterleitung der Kräfte auch im Diaphragmabereich.

Eine Patientin nach einem Motorradunfall mit Thoraxasymmetrie und Rippenschmerz brachte Kary (2009) auf die Untersuchung des M. Transversus Thoracis (Der Muskel zieht vom posterioren Sternum zur zweiten bis sechsten Rippe). Mit einer ausführlichen anatomischen und physiologischen Recherche beschreibt er ihn unter anderem als Diaphragmenspanner im Stehen, im besonderen bei älteren Menschen (Kary begründet

dieses Untersuchungsergebnis durch den Ausgleich der reduzierten Thoraxelastizität im fortschreitenden Alter). Kompressionskräfte durch Traumata (im hier beschriebenen Fall sind es segmentale thorakale Rotationsdysfunktionen) verändern die Spannung der Körpergewebe. Diese Spannung kann auf andere Körperbereiche übertragen werden über den M. Longus Colli, M. Pectoralis Minor, M. Transversus Abdominis, das Diaphragma oder das Pericard. Durch die Crura des Zwerchfells wird die erworbene Dysfunktion, so Kary, über M. Quadratus lumborum und M. Iliopsoas zum Lendenbereich, Becken und der unteren Extremität weitergeleitet. Vier anschauliche Fallbeispiele unterstreichen seine Gedankengänge. In seiner Untersuchung und seinen Behandlungen spielt übrigens auch das Diaphragma eine wesentliche Rolle.

Georges Finet und Christian Williame, zwei belgische Osteopathen, leisteten drei Jahre Forschungsarbeit über die physiologische Mobilität der Viscera bedingt durch die Zwerchfellbewegung. Zudem konnten sie bei Gastralgie, Diarrhoe und Opstipation spezifische Einschränkungen der Mobilität feststellen (Finet & Williame, 2007). Spannung und Mobilität des Diaphragmas und der Viscera beeinflussen sich also gegenseitig.

Eine andere Möglichkeit, Zusammenhänge vom Diaphragma zum restlichen Körper darzustellen, ist der Blick auf die Embryologie. Die Zwerchfellanlage entwickelt sich auf Höhe der cervikalen Somiten. Durch den Sog von Lunge und Herz wird das Zwerchfell nach caudal gezogen, der N. Phrenicus aus den Segmenten C3-C5 erinnert an den Descensus. Der Ösophagus, die Aorta, und Vena Cava Inferior, die laterale und anteriore Thoraxwand mit der Pleura, dem Peritoneum und das dorsale Mesenterium des Darms seien hier als selbige Gewebsanlage erwähnt (Helsmoortel, 2002, S.74; Meert, 2007, S.120).

Folgende philosophische Gedanken von Jaap van der Wal sollen zum Bezug der embryonalen Entwicklung zur jetzigen Form und Funktion des Diaphragmas inspirieren: "Im Embryo sieht man fortwährend, dass eins sich gliedert, und in zwei polarisiert... Zwischen den Polen atmet das Leben, in der sogenannten Mitte haben wir Rhythmus, Gleichgewicht, Gesundheit, Heilung, Prozess, in der Mitte atmet das Leben" (Schmidt & Liem, 2012).

2.3.2 Dysfunktion und Untersuchung vom Diaphragma

Die Thoraxform sowie die Körperhaltung beeinflussen den Zwerchfellstand. Zunahme des Füllungsdrucks im Bauchraum (Meteorismus, Schwangerschaft) treiben das Zwerchfell nach cranial in die Expirationsposition (Zwerchfellohochstand), wie es in Abb.7 ersichtlich ist (Leonhardt et al, 1987, S.306). Ein Zwerchfelltiefstand wie bei der Inspirationsposition wird z.B. durch Senkung, Verklebung der Bauchorgane, Bauchmuskelschwäche oder schnelle Gewichtsabnahme ausgelöst (Meert, 2007, S.141).

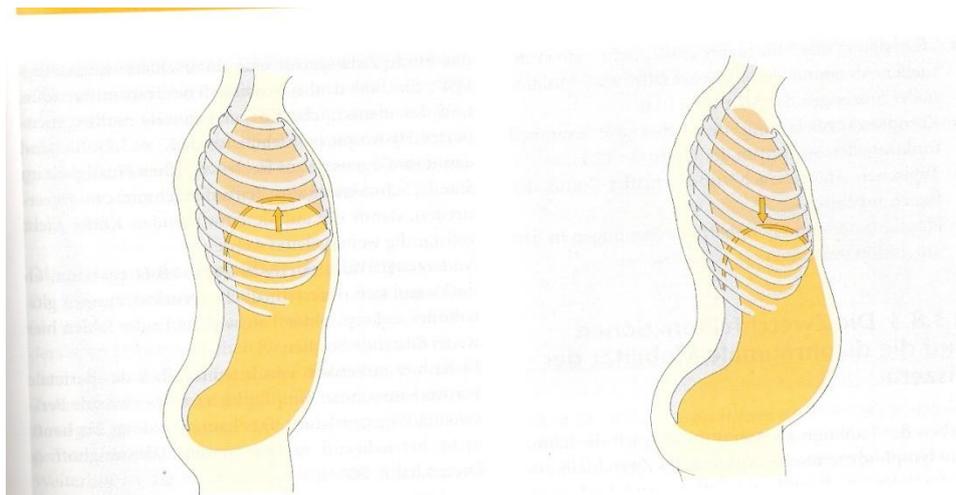


Abb.7: Fixierter Zwerchfellhochstand bzw. Fixierter Zwerchfelltiefstand (Meert, 2007, S.141).

Defeo & Hicks (1993) beschreiben die Haltungsanpassung des Rumpfes durch Schwerkraftwirkung und Mikrotraumen vorrangig an den Übergangsbereichen der Wirbelsäule (craniocervikal, cervikothorakal, thorakolumbal und lumbosacral), was mit den transversalen Diaphragmen (Tentorium Cerebelli, Thoracic Outlet, abdominales Diaphragma, Beckendiaphragma) örtlich übereinstimmt. Haltungsanpassungen an diesen Übergängen führen zu veränderten Ansätzen für die transversalen Diaphragmen und dadurch kommt es zu Restriktionen. Restriktionen bei transversalen Verbindungen haben laut Pope (2003) eine große gesamtkörperliche Auswirkung auf die Gesundheit und Anpassungsfähigkeit des Körpers bei eventuellen Traumata. Die Spannung wird zwischen den Diaphragmen in Rotation oder Lateralflexion am besten weitergeleitet. Gesund (und kompensationsfähig auf eventuelle Störeinflüsse) gilt der Körper dann, wenn sich die dominante Rotationsrichtung in den darauffolgenden Diaphragmen abwechselt (siehe Abb.8). Diese These stützt sich auf Untersuchungen aus dem Jahre 1979 von J. Gordon Zink (Glover et al, 2006; Pope, 2003), die durch Sniders Untersuchungen bekräftigt werden konnte (2008).

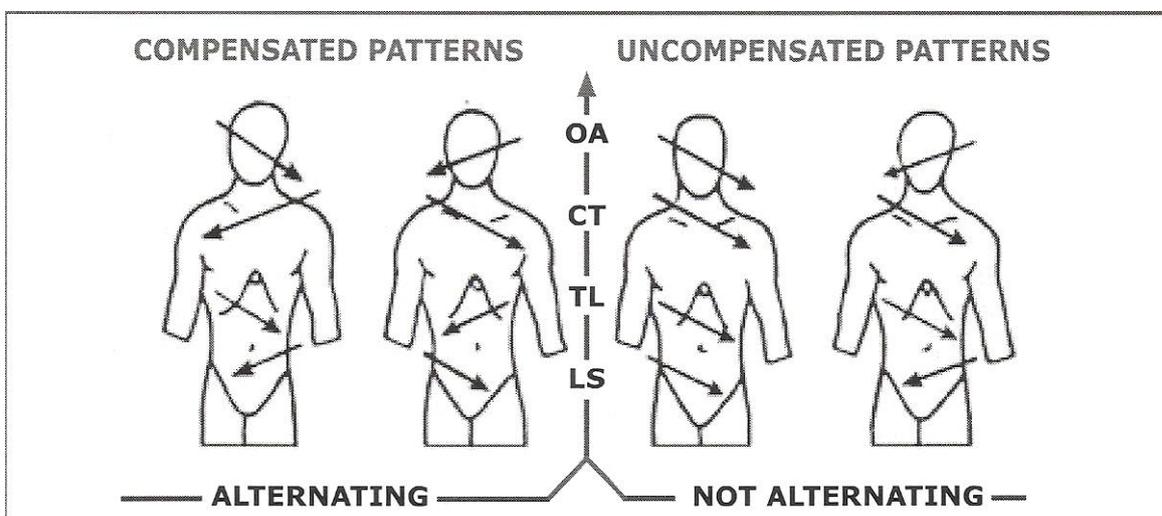


Abb.8: Kompensationsfähige und nicht kompensationsfähige Muster (Pope, 2003)

Ward beschreibt den Zusammenhang der Körpermuster treffend: "For every tightness, there is a three-dimensionally related looseness. Commonly the looseness is in exactly the opposite direction from the tightness." (Danto, 2003).

In einem Behandlungsüberblick (Kuchera, 2007), einer Literaturzusammenschau über osteopathische Behandlung von schwangeren Frauen (Lavelle, 2012) und einem Fallbeispiel mit Fazialslähmung (Lancaster & Crow, 2006) wird das Verringern der Restriktionen an den vier Diaphragmen als erfolgsversprechend beschrieben. Auch im Unterricht von Georges Finet, Christian Fossum, Bernard Ligner, und Dr. Erich Mayer-Fally an der WSO waren die Wichtigkeit, die Testung, die Behandlung und das Zusammenspiel der Diaphragmen Thema.

In osteopathischen Fachbüchern (Zusammenfassung aus Helsmoortel, 2002, S.91; Liem, 2005, S.485; Meert, 2007, S.324; Münch, 2011, S.108; Paoletti, 2001, S.203) werden folgende Tests beschrieben. Die Testausgangsstellung für den Patienten ist entweder der Sitz oder die Rückenlage (die Tests beantworten die nachfolgend angereicherten Fragen):

- statische Tests ohne Berücksichtigung der Atmung:
Steht das Diaphragma normal, hoch, tief oder in Torsion?
Ist der Tonus normoton, erhöht, erniedrigt, asymmetrisch?,
- respiratorische Tests mit Berücksichtigung der Atmung:
Senkt sich das Diaphragma symmetrisch/asymmetrisch?
Entspricht die Diaphragmabewegung der physiologischen Atembewegung?,
- dynamische Tests:
Reagieren die Rippen bzw. das Diaphragma auf eine vom Untersucher induzierte Bewegung wie eine Torsion, eine Translation oder eine Pumpbewegung von anterior prompt, zeitversetzt, gar nicht oder nur in eine bestimmte Richtung?
- Inhibitionstests:
Reagieren die Rippen bzw. das Diaphragma auf manuelles Anheben des gesamten peritonealen Sacks nach cranial prompt, zeitversetzt, gar nicht oder nur in eine bestimmte Richtung? Entsteht dadurch mehr Mobilität, wird angenommen, dass die Thoraxstrukturen eine intraperitoneale Dysfunktion kompensieren.
- Provokationstests:
Reagieren die Rippen bzw. das Diaphragma auf Zug des gesamten peritonealen Sacks nach caudal prompt, zeitversetzt, gar nicht oder nur in eine bestimmte Richtung? Entsteht dadurch mehr Mobilität, wird angenommen, dass die peritonealen Strukturen eine thorakale Dysfunktion kompensieren.

Die Schwierigkeit von dynamischen Tests liegt beim feinen Impuls in die zu testende Richtung. "Der Untersucher folgt der induzierten Bewegung mit passiver Aufmerksamkeit bis zu ihrem Endpunkt. Außer dem anfänglichen Impuls wird die Bewegung nicht weiter vom

Behandler gelenkt, sondern nur begleitet. Der Untersucher nimmt wahr, ob die Bewegung zugelassen wird oder eingeschränkt ist, mit welcher Qualität, in welche Richtung und auf welche Art welche Struktur auf den feinen Bewegungsimpuls reagiert" (nach Liem, 2006, S.230).

Trotz intensiver Recherchen konnten für die Zwerchfelltestung keine Studien bezüglich ihrer psychometrischen Eigenschaften gefunden werden. Aufgrund der fehlenden Studien kann aktuell über die Reliabilität der Diaphragmentests keine Aussage getroffen werden. Es fehlt außerdem eine Beschreibung einer standardisierten Testausführung. Dadurch ist nicht klar, in welcher Palpationstiefe die Tests stattfinden. Es ist unklar, ob beim Testen des Diaphragmas mit den oben angeführten Tests ebenso die Fascia Superficialis, die Intercostalmuskulatur, die Rippen, die Fascia endothoracica, die Pleura, die Eigenspannung der Leber, des Magens, der Milz und/oder der Niere untersucht werden. Die Palpationstiefe, also die Frage, welche Struktur getestet wird, ist demnach eine Interpretation des Untersuchers.

Laut Tensegritymodell stehen einzelnen Strukturen miteinander in Verbindung und haben so auf einander Einfluss. Ist z.B. die Pleura vermehrt gespannt, leitet sich die Spannung auf das Diaphragma und zur Eigenspannung der Leber weiter.

Interessant in diesem Zusammenhang ist hier die Frage nach dem Ursprungsort der Spannung. Kann dieser Ort mit Tests gefunden, also zwischen den Strukturen differenzieren?

Wie objektiv oder subjektiv diese Testaussagen sind, kann mit einer methodischen Studie beantwortet werden.

3. Empirischer Teil

Nach der anfänglichen Beschreibung des zu untersuchenden Tests werden die Forschungsfragen mit ihren Hypothesen formuliert. Die methodische Herangehensweise zur Beantwortung der Fragestellungen wird in den darauffolgenden Kapiteln beschrieben. Abschließend werden die verwendeten statistischen Tests zur Beantwortung der Fragestellung und zur weiteren Ursachenforschung erklärt.

3.1 Definition der Testung der abdominalen Diaphragmaspannung

In dieser Studie wurde der dynamische Test mit induzierter Translation in der Frontalebene wegen guter Beschreibung in der Fachliteratur (Helsmoortel, 2002, S.92; Meert, 2007, S.324; Münch, 2011, S.108; Paoletti, 2001, S.203), Vorstellung im Unterricht von Bernard Ligner, Dr. Erich Mayer-Fally bzw. Christian Fossum und aus persönlicher Vorliebe gewählt. Spannungsdifferenzen lassen sich mit diesem Test im Seitenvergleich schnell feststellen. Angelehnt an die von Meert (2007, S.324) beschriebene Untersuchung wird die Testung der abdominalen Diaphragmaspannung wie folgt untersucht (siehe Abb.9):

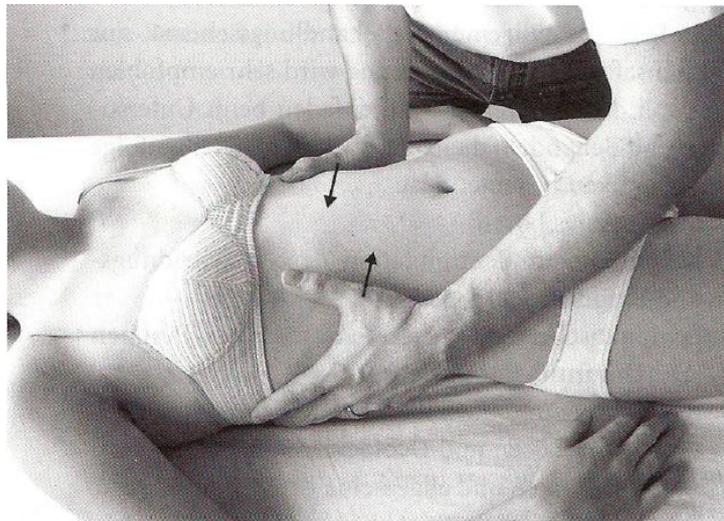


Abb.9: Test der abdominalen Diaphragmaspannung (Meert, 2007, S.324)

Der Untersucher legt seine Hände beidseits auf den Rippenbogen des Probanden von lateral kommend. Der Diaphragmaansatz darunterliegend wird visualisiert. Der Brustkorb des Probanden wird passiv und leicht von der einen zur anderen Seite transversal verschoben. Beurteilt wird der Bewegungsbeginn, es findet also keine Maximalbewegung statt. Die Angleichung der Bewegungsamplitude wird im Punkt 3.5.3. "Einschulung der Osteopathen" näher erläutert. Der Untersucher beurteilt, ob die Spannung im Gesamteindruck beidseits gleich (=G), rechts (=R) oder links (=L) unterschiedlich ist. Es wird die Seite angegeben, die besser bewegt und Bewegung besser durchlässt.

3.2 Forschungsfragen und Hypothesen

Ziel der Studie ist die Beantwortung folgender Forschungsfragen:

1. Kommen zwei Osteopathen, die am selben Probanden (mit und ohne Beschwerden) einen Test am abdominalen Diaphragma in Translation ausführen, zum gleichen Ergebnis (Interrater Reliabilität)?
2. Kommt ein und derselbe Osteopath, der am selben Probanden (mit und ohne Beschwerden) mehrmals einen Test am abdominalen Diaphragma in Translation ausführt, zum gleichen Ergebnis (Test-Retest Reliabilität)?
3. Stimmen die Testergebnisse und die Schmerzseite bei Patienten mit chronischen Schmerzen im lumbalen, thorakolumbalen und thorakalen Bereich überein?

3.2.1 Hypothese 1

H0: Führen zwei Osteopathen innerhalb von zwei Stunden am selben Probanden einen Test am abdominalen Diaphragma in Translation aus, liegt die Übereinstimmung ihrer Ergebnisse im zufälligen Bereich ($p > 0,05$).

H1: Führen zwei Osteopathen innerhalb von zwei Stunden am selben Probanden einen Test am abdominalen Diaphragma in Translation aus, ist die Übereinstimmung ihrer Ergebnisse statistisch signifikant ($p > 0,05$).

3.2.2 Hypothese 2

H0: Führt ein Osteopath innerhalb von zwei Stunden am selben Probanden einen Test am abdominalen Diaphragma in Translation aus, liegt die Übereinstimmung seiner Ergebnisse im zufälligen Bereich ($p > 0,05$).

H1: Führt ein Osteopath innerhalb von zwei Stunden am selben Probanden einen Test am abdominalen Diaphragma in Translation aus, ist die Übereinstimmung seiner Ergebnisse statistisch signifikant ($p > 0,05$).

3.2.3 Hypothese 3

H0: Die Übereinstimmung der Ergebnisse des Diaphragmatests und der Schmerzseite bei Patienten mit chronischen Schmerzen im lumbalen, thorakolumbalen oder thorakalen Bereich liegt im zufälligen Bereich ($p > 0,05$).

H1: Die Übereinstimmung der Ergebnisse des Diaphragmatests und der Schmerzseite bei Patienten mit chronischen Schmerzen im lumbalen, thorakolumbalen oder thorakalen Bereich ist statistisch signifikant ($p < 0,05$).

3.3 Forschungsdesign

Die Hypothesen werden mittels einer Interrater- und Test Retest Reliabilitätsstudie überprüft. Generell zeigen Reliabilitätsstudien in der manuellen Medizin und in der Osteopathie eine geringe Zuverlässigkeit (Haneline & Young, 2009; Olivo et al, 2008; Patijn, 2004; Stovall & Kumar, 2010). Daher und wegen der unterschiedlichen Vorgangsweise hat die International Federation for Manual/Musculoskeletal Medicine (FIMM) sieben goldene Regeln als Leitfaden zur besseren Validität vorgeschlagen (Patijn, 2004):

1. **Für den Überblick der gesamten Studie ist ein Untersuchungsleiter vonnöten.**
In dieser Studie übernimmt diese Aufgabe die Autorin.
2. **Eine Trainingsperiode vor der methodologischen Studie hilft bei der Auswahl des Tests, der Bestimmung der Probandenzahl und der Diskussion über die Relevanz des Tests in der Praxis.** Auf die Handhabung dieser Forderung wird in Punkt "3.5.1. Probedurchlauf" näher eingegangen.
3. **Eine Übereinstimmung von 80% im Probedurchlauf erhöht die Wahrscheinlichkeit auf eine reliable Studie.** Zusätzlich zum Probedurchlauf und zur Einschulung der Untersucher wurde dieser Prämisse im Selbsttest der Autorin während der Praxisarbeit nachgegangen. Die Eigenwahrnehmung des Patienten und das palpierende Ergebnis der Autorin beim Spannungstest am abdominalen Diaphragma wurden verglichen, was eine 100% tige Übereinstimmung ergab. Diese Palpationsübung wurde ein Monat vor der Studie den beiden Untersuchern für fünf Patienten pro Tag empfohlen.
4. **Die Blindierung sollte sowohl für den Probanden, als auch für den Untersucher gegeben und dokumentiert sein.** Durch das Abdecken der oberen Körperhälfte des Probanden mit einem Leintuch konnte weder der Proband den Untersucher sehen, noch der Untersucher den Probanden. (siehe 3.5.1.).
5. **Die Probandenpopulation ist gegenüber der Gesamtpopulation abzugrenzen.** Die Probanden wurden einerseits aus dem Patientenpool der Autorin, andererseits durch Anfragen bei Verwandten, Freunden und Bekannten ausgewählt. Das Telefonieren sowie Aussenden von Informationsmaterial an Kollegen und Ärzte (siehe Anhang "Ärzteinformation") erbrachte keinen Zugewinn an Probanden. Die Probanden wurden über den Ablauf der Studie mittels eines Informationsblattes aufgeklärt (siehe Anhang "Probandeninformation"). Die Daten der Probanden wurden in Form eines Fragebogens (siehe Anhang "Fragebogen für die Probanden") ermittelt und in die jeweiligen Gruppen eingeteilt.
6. **Der Studienablauf, wie die Definition des untersuchten Tests und die Methoden zur Berechnung der Testergebnisse sollten transparent sein,** siehe "3.5. und 3.6." Olivo et al (2008) bemerkt die zusätzlich schlechte Vergleichbarkeit, wenn jeder

Autor für seine Studie eine eigene Skalierung erfindet. Mit den Kappa-Indices, den χ^2 -Tests und Exakte Tests nach Fisher ist dies nicht gegeben.

7. Die Ergebnisse sind in lesbaren Tabellen abzubilden.

3.4 Stichprobenbeschreibung

3.4.1 Untersucher (Osteopathen)

Die gesichteten Reliabilitätsstudien arbeiten meist mit zwei Untersuchern (Clare, Adams & Maher, 2005; Degenhardt, Johnson & Snider, 2010; Halma et al 2008; Kilpikoski et al, 2002; Moran & Gibbons, 2001; Snider et al, 2008; Sommerfeld, Kaider & Klein, 2004) - Ausnahmen bilden die Studie Sergueef, Greer, Nelson & Glonek (2011) mit 734, von Bengaard, Bogue & Crow (2012) mit 151 Prüfern, von Parodi, Chauvigny de Plot, Rickards & Renard (2009) mit zwölf und die Masterthese von Rittler (2010) mit sechs Untersuchern.

Patijn (2004) sieht keinen statistischen Grund für mehr als zwei Untersucher. Die Tagesverfassung wird allerdings bei zwei Untersuchern mehr gewichtet und verringert die Aussagekraft der Studie (Rittler, 2010). Das Problem der Vergleichbarkeit bei erfahrenen Untersuchern liegt nach Patijn (2004) in der nach mehreren Berufsjahren eigens entwickelten Art der Ausführung und Interpretation des Tests. Sergueef et al (2011) hingegen stellte fest, dass die Daten von erfahrenen Untersuchern weniger gestreut sind.

Für diese Studie sind sowohl für die Interrater- als auch für die Test-Retest Reliabilität zwei Osteopathen geplant. Die ausgewählten Osteopathen haben die gleiche Anzahl der Tätigkeitsjahre (Abschlussprüfung an der WSO 2006), weisen die gleiche Grundausbildung aus (Ausbildung zum diplomierten Physiotherapeuten im AKH Wien 1994-1998) und haben ähnliche Zusatzausbildungen abgeschlossen (Spiraldynamik, Biodynamik von Jim Jealous). Zudem arbeiten beide Osteopathen bimanuell, ihre Schreibtätigkeit wird rechts ausgeführt.

3.4.2 Probanden

Für die Teilnahme an der Studie werden Personen mit und ohne Schmerzen im lumbalen, thorakolumbalen und thorakalen Bereich, die die unten angeführten Ein- und Ausschlusskriterien erfüllen, gesucht.

Laut Fritz & Wainner (2001) soll die Studienpopulation die klinische Praxis repräsentieren wobei die Untersuchungsparameter von mild bis stark auftreten sollten (Fritz & Wainner, 2001).

Es soll eine gleichmäßige Geschlechterverteilung angestrebt werden. Die Studienpopulation repräsentiert die klinische Praxis, wobei die Untersuchungsparameter von mild bis stark auftreten sollten (Fritz & Wainner, 2001).

Dass schweres Körpergewicht die Palpation kompliziere, konnte von Degenhardt (2010) und Wolke (2009) durch die Einbeziehung des Bodymaßindex nicht bestätigt werden.

Halma et al (2008) teilte seine Probanden für die Testung der "cranial strain pattern" in eine Asthma-, eine Kopfwehgruppe und eine gesunde Kontrollgruppe ein. Bei seiner Studie stellte er bei der gesunden Kontrollgruppe ($\kappa= 0,82$) im Vergleich zu den Symptomgruppen ($\kappa= 0,67$ bei Kopfwehgruppe, $\kappa= 0,52$) eine nahezu perfekte Test-Retest Reliabilität fest. Er begründet dies mit dem stabileren gesundheitlichen Zustand. Die schlechtere Reliabilität bei den Symptomgruppen im Vergleich zur gesunden Kontrollgruppe begründet Halma mit dem periodischen Auftreten der Schmerzen. Allerdings waren die drei Durchgänge der Untersuchung in einem Zeitraum von nur 45-60min abgeschlossen, was die Periodizität der Symptomgruppen in Frage stellt.

Im Gegensatz dazu sind bei chronischen Schmerzen im thorakolumbalen Bereich Gewebs- und Bewegungsänderungen ohne Periodizität spürbar, das Gewebe ist dort weniger beweglich und zieht zur betroffenen Seite hin (Kuchera, 2007; Snider et al, 2008). Diese Beobachtung könnte ein Grund sein, dass Gewebsunterschiede im Seitenvergleich bei chronischen Schmerzpatienten im beschriebenen Bereich besser feststellbar sind und dadurch eine bessere Vergleichbarkeit in der Testung erzielt werden kann.

Mit Teilung der Probandengruppe in Gesunde und Schmerzpatienten soll geprüft werden, ob die Gewebsspannung zuverlässiger bei chronischen Schmerzpatienten zu testen ist.

3.4.3 Einschlusskriterien

Erwünscht sind zwanzig Personen zwischen 18 und 75 Lebensjahren mit einem subjektiv stabilen gesundheitlichen Zustand. Zudem werden zwanzig Patienten gesucht, die Schmerzen im lumbalen, thorakolumbalen oder thorakalen Bereich mindestens fünf Tage in der Woche seit mindestens drei Monaten aufweisen (Definition für Chronizität nach Snider et al, 2008).

3.4.4 Ausschlusskriterien

Aufgrund der Kontraindikation der Untersuchungstechnik sind Personen mit akutem Geschehen im Thorax und Abdomen (Pneumonie, Pneumothorax, Lungenembolie, Frakturen) ausgeschlossen. Ferner dürfen schwangere Frauen im letzten Trimenon wegen erhöhter Wiedererkennungswahrscheinlichkeit bei der Test-Retest Reliabilitätsüberprüfung nicht teilnehmen. Die Ausschlusskriterien wurden telefonisch abgefragt.

3.4.5 Demographische Angaben

Insgesamt nahmen an der Studie 43 Probanden im Alter von 30-75 Jahren (M = 56 Jahre; SD = 13,7 Jahre; Median: 60 Jahre) teil, davon 24 Frauen und 19 Männer (siehe Abb.11 und

Anhang). Der BMI der Probanden betrug im Durchschnitt $M= 24,4 \text{ kg/m}^2$; $SD= 3,0 \text{ kg/m}^2$ (19,8 bis $30,7 \text{ kg/m}^2$, Median = $24,2 \text{ kg/m}^2$).

Neben 22 Probanden ohne Schmerzsymptomatik umfasste die Stichprobe 21 Schmerzpatienten (neun Probanden mit Schmerzen links, fünf rechts, sieben beidseits). Der Schmerz war bei fünf Probanden im thorakalen Bereich, bei zwei im thorakolumbalen Übergang und bei vierzehn im lumbalen Bereich.

3.5 Studienablauf

3.5.1 Probedurchlauf

Für den Probedurchlauf, der am 05. 09. 2012 am späteren Studienort in der Praxis der Autorin in Gmunden stattfand, waren fünf Probanden (eine Probandin wurde zwecks Feedbackkontrolle von der Autorin verkörpert) und zwei Tester anwesend. Die Untersucher bei der Proberunde sind mit denen der Studie nicht ident. Es galt einen Untersuchungsablauf durchzuführen, der neben der Optimierung der Stichprobengröße, des Managements und der Dokumentation folgende Bereiche entscheidet:

1. Handhaltung der Probanden:

Darauf wurde in der Beschreibung des Tests in der Literatur nicht eingegangen. Der Proband legt seine Arme neben seinem Körper auf die Behandlungsliege. Andere Armstellungen beeinflussen die Diaphragmaspannung zu sehr und könnten zudem Schmerzen während der Testung hervorrufen (Armkreuz vor dem Körper, verschränkte Armhaltung über Kopf).

2. Knierolle:

Jeder Proband bekommt eine Knierolle. Schmerzpatienten haben aus Erfahrung vermehrt Schmerzen in der ausgestreckten Rückenlage, die nach einiger Zeit nachlassen. Diese Situation würde die Reproduzierbarkeit des Tests minimieren.

3. Griffhaltung:

Nach einer Feedbackrunde mit den fünf Probanden mit dem Ergebnis, dass die Griffausführung der Untersucher unterschiedlich gespürt wurde, wurde deutlich, dass die Untersucher eine einheitliche und klare Einschulung benötigen. Das Ergebnis der Testung (siehe Tab.1), dass wenig Übereinstimmung zeigte, war ein weiterer Hinweis für die Wichtigkeit dieser Einschulung. Im Rahmen der Einschulung muss die Testausführung mit Betonung des flächigen Griffs, der den Eigenschwung des Brustkorbs des Probanden in die induzierte Bewegung mitnimmt, geklärt werden. Der Test wird in seiner Stärke, Schnelligkeit, Griffweise und Art der Ausführung angeglichen. Klein & Sommerfeld (2004, S.104) konnten mit diesen Parametern eine Beeinflussung des viskoelastischen Verhaltens im Bindegewebe feststellen. Weiters

soll die Testausführung bei zwei Personen, die klares Feedback geben können, angeglichen werden. Im Feedback der Probanden ist die Palpationstiefe der Untersuchung auch eine Fragestellung.

4. **Atmung des Probanden:**

Nach theoretischen Überlegungen wäre der Test durch die maximale Ausdehnung des Diaphragmas in der Inspirationsstellung des Probanden am sinnvollsten. Jedoch ist dies durch Nervosität bzw. durch das Luft anhalten der Probanden schwierig durchführbar und endet in einer vermehrten Grundspannung, dies würde eine Beurteilung insgesamt wieder erschweren. Daher wurde die Instruktion "entspanntes Liegen mit gleichmäßigem Atmen" für die Probanden gewählt.

5. **Blindierung:**

Die Blindierung erfolgt mit einem Leintuch, das wie ein Vorhang zwischen zwei Garderobenständer gespannt wird (siehe Abb.10). Diese werden auf Höhe des oberen Rippenbereichs des Probanden beidseitig platziert. Der Proband wird ab den Beckenknochen mit einem Leintuch zugedeckt. So ist der Bauchbereich zwischen den unteren Rippen und Beckenknochen, wie auch die Unterarme des Probanden für den Untersucher sichtbar. Die Hände des Osteopathen haben einen Zugang zum Untersuchungsbereich. (Leider wurde das Leintuch für den Beinbereich der Probanden während der Hauptstudie vergessen.)

Das Verbinden der Augen des Osteopathen wurde auch in Erwägung gezogen. Dies erschwert jedoch die Beurteilung des Tests, da nicht nur die Palpation, sondern auch die Augen die Spannung beurteilen (Krause, 2008, S.13). Außerdem entspricht der Test ohne Verbinden der Augen vermehrt der Praxisrealität. Dass sich die Probanden nicht entkleiden müssen, bringt eine große Zeitersparnis mit sich.



Abb.10: Blindierung für Proband und Untersucher

6. Organsiation:

Vier Assistenten werden für einen reibungsfreien Ablauf benötigt. Zwei Assistenten werden eingeschult, die Ergebnisse der Tester auf eine vorgefertigte Tabelle einzutragen. Die Assistenten werden unterwiesen, bei einem sicheren Palpationsergebnis das Ergebnis einzukreisen. Eine Studie belegt, dass die Übereinstimmung bei Sicherheit im Ergebnis größer ist (Krause, 2008, S.8). Als sicher wurde das Testergebnis definiert, wenn der Untersucher eindeutig den Spannungszustand feststellen kann (jede induzierte Bewegung ergibt dasselbe Ergebnis). Der Spannungstest am Diaphragma wird zweimal auf jeder Seite durchgeführt. Ein Buffet während der Untersuchung würde die Diaphragmaspannung der Probanden verändern. Das Buffet wird NACH der Untersuchung eröffnet.

Für die Testung von fünf Probanden mit zwei Durchgängen, also für zehn Untersuchungen, wurden vierzehn Minuten benötigt. Für 60 Untersuchungen wären demnach 90 min ausreichend.

In Tab.1 sind die Ergebnisse der Probedurchführung ersichtlich:

Tab.1: Ergebnisse der Probedurchführung 05.09.2012

Raum A				Raum B			
Reihenfolge	Proband	Tester 1	Tester 2	Reihenfolge	Proband	Tester 2	Tester 1
1 A	G	L	G	1 B	T	L	G
2 A	M	L	R	2 B	S	G	L
3 A	G	L	L	3 B	T	L	G
4 A	B	G	G	4 B	B	L	G
5 A	M	G	L	5 B	S	G	G

Legende: L = links, R = rechts, G = gleich, **Fett** = sicheres Ergebnis

3.5.2 Einschulung der Osteopathen

Kilpikovsky et al (2002), Clare et al (2005) und Bengaard et al (2012) geben im Gegensatz zu Parodi et al (2009) bessere Ergebnisse mit gemeinsamen Training der Untersucher vor der Studie an. Degenhardt et al (2010) geht mit seiner Folgestudie von 2005 sogar so weit, dass er nach zwei Monaten Trainingsmodus bei Uneinigkeit seiner beiden Tester nochmals eine Nachschulung einfügt, um die darauffolgende viermonatige Studiendauer zu optimieren. Die Interreliabilität konnte so deutlich verbessert werden (von $\kappa \geq 0,24$ auf $\kappa \geq 0,40$; $p < 0,001$).

Aus örtlichen Gegebenheiten lässt sich die geplante Einschulung nicht ein Monat vor der Studie durchführen. Stattdessen wird die genaue Testanleitung per Mail mit Bild (siehe 3.1.) verschickt und wie in 3.2. erwähnt die Anwendung jeden Arbeitstag bei fünf Patienten mit Patientenfeedback für einen Monat empfohlen.

Am Studientag findet eine Einschulung (wie unter Punkt 3.5.1. beschrieben) für die Osteopathen statt. Der Untersuchungsablauf wird erläutert. Da beide Untersucher ein dominantes rechtes Auge aufweisen, kann die Idee von Bengaard et al (2012) aufgegriffen werden: Der Untersucher sollte sich rechts neben dem Behandlungsbett stellen.

Es wurde von der Autorin die Verwendung eines Algometers in Erwägung gezogen. Dieses Gerät kann den ausgeübten Fingerdruck messen und wird z.B. in der Akupressurbehandlung verwendet. Für eine flächige Palpation gibt ein Algometer aber keine sinnvolle Orientierung. Die Palpationstiefe sollte mit der Intention des Diaphragmaansatz zu spüren bestimmt werden, dies betrifft auch den Winkel der Hände zum Rippenbogen.

3.5.3 Randomisierung

Die Untersuchungsreihenfolge von Untersucher und Probanden wurde von der Autorin unter Beisein eines Zeugen per Los entschieden. Zusätzlich wurden von den 43 Probanden zwanzig per Los ausgewählt, die im zweiten Testdurchgang bei der Test-Retest Untersuchung teilnahmen (siehe Abb.11). Die Untersuchungsreihenfolge kann Aufschluss über eventuelle Behandlungseffekte bei aufeinanderfolgenden Untersuchungen geben.

3.5.4 Untersuchungsablauf

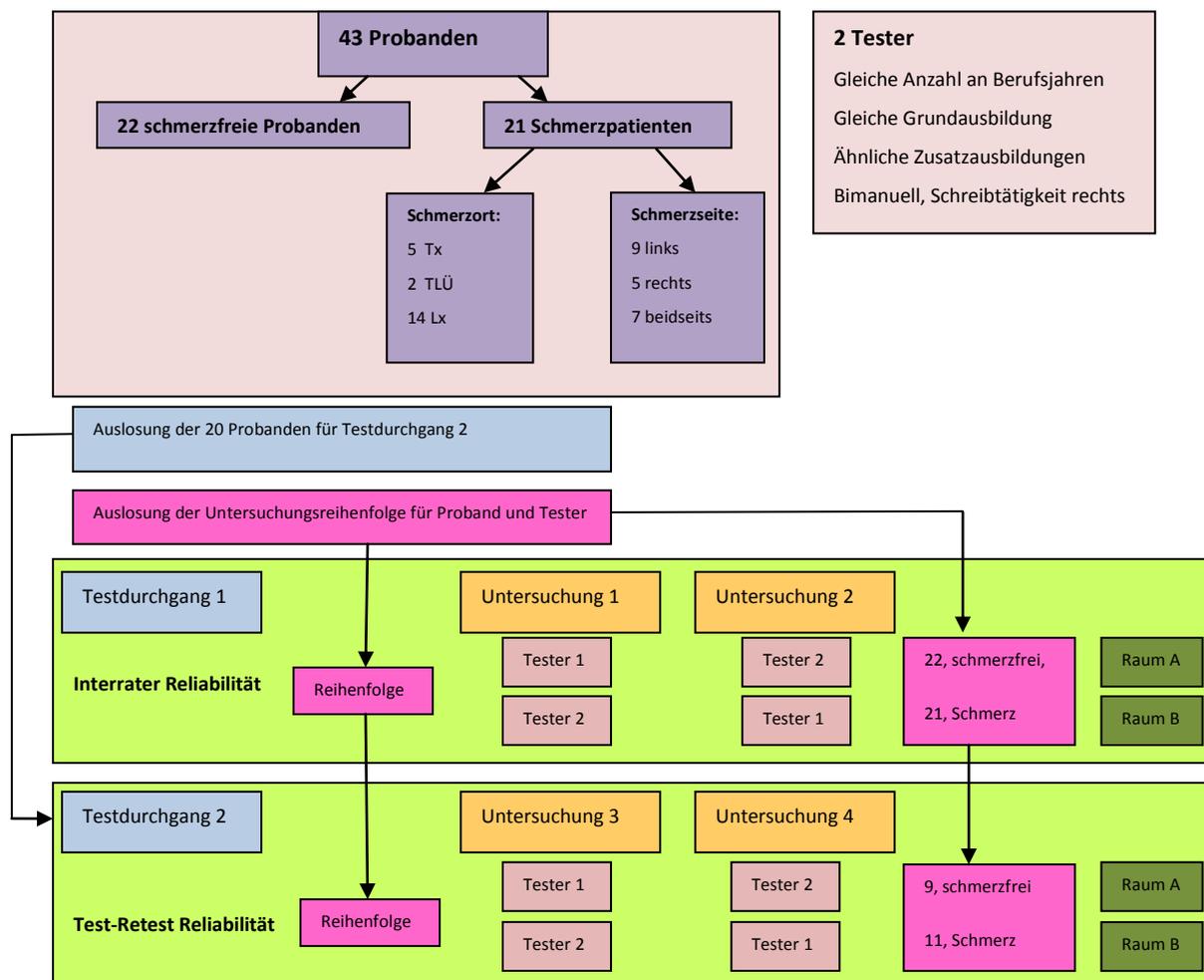


Abb. 11: Schematischer Überblick zum Untersuchungsablauf

Abb. 11 bietet einen schematischen Überblick vom Untersuchungsablauf am 19. 10. 2012. Von den 43 Probanden (22 schmerzfreie Probanden und 21 Schmerzpatienten), die beim ersten Testdurchgang in geloster Reihenfolge zur Interrater Reliabilität untersucht werden, nehmen zwanzig (neun schmerzfreie Probanden und elf Schmerzpatienten) beim zweiten Durchgang zur Test-Retest Überprüfung teil.

Zwei Untersuchungsräume ermöglichen ein zeitgleiches Testen von zwei Probanden. Der Untersucher 1 testet als Erster im Raum A und als Zweiter im Raum B, der Untersucher 2 testet als Erster im Raum B und als Zweiter im Raum A.

Die Untersuchungsliegen werden vor Studienbeginn auf die Höhe der Osteopathen eingestellt. Die Assistenten werden in ihre nachfolgend beschriebenen Aufgaben eingewiesen. Die Untersuchungsleiterin begrüßt die eintreffenden Probanden im Warteraum und überreicht den Fragebogen zum Ausfüllen (siehe Anhang "Fragebogen für die Probanden").

Die zugeteilte Numerierung der Probanden wird mit gut sichtbaren Aufkleber in Schlüsselbeinhöhe versehen. Der erste Assistent geleitet die Probanden in der gelosten

Reihenfolge zu Raum A, der zweite Assistent in Raum B. Der dritte Assistent blindiert die Probanden, gibt dem Osteopathen Bescheid, wann der Raum zu betreten ist und die Untersuchung starten kann. Ebenso notiert dieser Assistent die Ergebnisse der Osteopathen im Raum A in einer vorgefertigten Tabelle (siehe Anhang "Originale Eintragungen der Untersuchungsergebnisse"). Der vierte Assistent führt diese Aufgaben im Raum B durch.

Die Probanden werden angehalten, möglichst entspannt und ruhig zu liegen (Luft anhalten, Lachen,... verändert die Spannung erheblich). Wenn der Osteopath vom Assistenten in den Untersuchungsraum geleitet wird, liegt der Proband "blindiert" am Behandlungsbett.

Der Osteopath stellt sich auf Beckenhöhe rechts neben den Probanden und legt als Begrüßung sanft die Hand auf die Probandenarm (siehe Abb.12). Der Spannungstest am Diaphragma wird zweimal auf jeder Seite durchgeführt, anschließend kommuniziert der Osteopath das Ergebnis an den Assistenten.

Es wurden Handzeichen zwischen den Osteopathen und den Assistenten ausgemacht, um ein nonverbales Kommunizieren zu ermöglichen (Zeigen auf den Ort der Spannung: rechte Körperhälfte, linke Körperhälfte oder mittig; mit dem Daumen nach oben bei einem sicheren Palpationsergebnis; mit einer schwenkenden Handbewegung bei unsicherem Ergebnis).

Nach dem gesamten Untersuchungsablauf wird an die Probanden der Feedbackbogen ausgeteilt (siehe Anhang "Feedbackbogen Probanden") und das Buffet wird eröffnet.



Abb.12: Ausführung des Diaphragmentests bei der Studie

3.6 Statistische Datenverarbeitung und -auswertung

3.6.1 Grundlegendes

Für die Bewertung der Reliabilität wurde der Kappa-Index (κ -Index) nach Cohen (1960) verwendet. Der Kappa-Index ist ein Maß dafür, wieweit eine beobachtete Übereinstimmung zwischen zwei Untersuchern (im Fall der Interrater-Reliabilität), bzw. zwei Untersuchungen (im Fall der Test-Retest-Reliabilität) über jene Übereinstimmung hinausgeht, die durch Zufall zu erwarten wäre. Der Wert Null weist dabei auf eine rein zufällige Übereinstimmung hin, der Wert +1 auf völlige Übereinstimmung. Negative Werte, wie sie in dieser Studie ebenfalls zu beobachten sind, treten dann auf, wenn die Häufigkeit übereinstimmender Ergebnisse geringer ist, als bei zufälliger Übereinstimmung und können minimal den Wert -1 erreichen.

Landis & Koch (1977) schlagen folgende textliche Interpretation der κ -Indices vor:

Tab.2: Textliche Interpretation des Kappa-Index nach Landis & Koch 1977: 159-174

$\kappa < 0,20$	Poor	schwach
$0,20 < \kappa < 0,40$	Fair	leidlich
$0,40 < \kappa < 0,60$	Moderate	mittelmäßig
$0,60 < \kappa < 0,80$	Substantial	beträchtlich
$0,80 < \kappa < 1,00$	almost perfect	fast ideal

Laut Fjellner, Bexander, Faleji & Strender (1999, S.511-516) werden Werte von zumindest 0,4 als Indikator für eine akzeptable interindividuelle Verlässlichkeit angesehen.

3.6.2 Durchführung der statistischen Auswertung

Die handschriftlich von den Assistenten notierten Ergebnisse ("Originalia" siehe Anhang) der beiden Tester wurden in eine Microsoft® Excel®-Tabelle übertragen ("Reihenfolge gesamt" siehe Anhang). Die Ergebnisse wurden computerunterstützt unter Verwendung der Statistiksoftware R 2.14.1. (R Development Core Team, 2011) ausgewertet.

Neben den Kappa-Indices werden die Ergebnisse der z-Tests angeführt, womit die Nullhypothesen überprüft werden, dass der Kappa-Index gleich Null ist. Das Signifikanzniveau wurde dabei mit $\alpha=0,05$ gewählt.

Statistische Auswertung für die Hypothesen

Kappa-Indices nach Cohen wurden für folgende Aspekte berechnet:

- a) Interrater und Test-Retest Reliabilität,
- b) Übereinstimmungsort des Schmerzes (laut Patientenangabe) und dem der Testergebnisse (Berechnung für Tester 1 und Tester 2)

ad a) Interrater und Test-Retest Reliabilität

- 1) Die Bestimmung der Interrater Reliabilität erfolgte aus den 43 Ergebnissen jedes Testers aus den Untersuchungen 1 und 2 (Testdurchgang 1).
- 2) Zusätzlich wurde die Interrater Reliabilität aus den sechzehn Ergebnissen jedes Testers aus den Untersuchungen 1 und 2 (Testdurchgang 1) berechnet, bei denen zumindest einer angab, ein „sicheres“ Ergebnis vorgefunden zu haben.
- 3) Die Bestimmung der Test-Retest Reliabilität erfolgte getrennt für jeden der beiden Tester aus den jeweils zwanzig Ergebnissen der Untersuchungen 3 und 4 (Testdurchgang 2). Von einer zusätzlichen statistischen Auswertung der Ergebnisse, bei denen zumindest eine Untersuchung ein „sicheres“ Ergebnis war, wurde abgesehen, da dies nur bei vier Probanden der Fall war.

ad b) Übereinstimmungsort des Schmerzes (laut Patientenangabe) und den Testergebnissen
Analysiert wurde die Übereinstimmung der Körperseite (schmerzabgewandte Seite) 21 symptomatischen Patienten mit den Ergebnissen jedes einzelnen Testers aus dem Testdurchgang 1 (Untersuchung 1 bzw. 2). Es wird im Testergebnis die Körperseite mit höherer Beweglichkeit angegeben und sollte daher, nach der Studie von Snider (2008), der schmerzabgewandten Körperseite entsprechen.

Statistische Auswertung für die Ursachenforschung:

Kappa-Indices nach Cohen wurden für folgenden Aspekt berechnet:

Mögliche Behandlungseffekte

Durch Bestimmung der Interrater Reliabilität aus den 22 Ergebnissen jedes Testers, bei denen Tester 1 in Untersuchung 1 und Tester 2 in Untersuchung 2 testete, sowie den 21 Ergebnissen, bei denen Tester 2 in Untersuchung 1 und Tester 1 in Untersuchung 2 testete, wurde festgestellt, ob sich ein beobachteter Unterschied in der Durchführung des Tests, der möglicherweise Behandlungseffekte mit sich bringen könnte, auf die Interrater Reliabilität auswirkt.

Für weiterführende statistische Untersuchungen wurden **χ^2 -Tests** (Chi-Quadrat-Tests) durchgeführt, wobei das Signifikanzniveau ebenfalls mit $\alpha=0,05$ gewählt wurde.

Diese Tests wurden dafür verwendet, um

- a) Unterschiede zwischen den beiden Testern hinsichtlich der Häufigkeitsverteilung von Übereinstimmungen und Nicht-Übereinstimmungen bei den Untersuchungen 3 und 4 (Testdurchgang 2 für die Bestimmung der Test-Retest Reliabilität) festzustellen und

b) um zu beurteilen, ob es einen Unterschied zwischen den beiden Testern in der Häufigkeitsverteilung der nominalskalierten Untersuchungsergebnisse („Rechts“/„Links“/„Gleich“) gibt. Dadurch könnten individuelle Präferenzen in den Untersuchungsergebnissen der beiden Tester sichtbar gemacht werden. Datengrundlage waren dafür die 43 Ergebnisse, die während des ersten Testdurchgangs gewonnen wurden. Obwohl eine Erhöhung der Fallzahl durch Miteinbeziehung der Ergebnisse des zweiten Testdurchgangs (Bestimmung der Test-Retest Reliabilität) möglich gewesen wäre, wurde davon Abstand genommen, da diese Daten an Probanden gewonnen wurden, die bereits im Testdurchgang 1 untersucht worden sind. Gesehen den Fall einer Wiederholbarkeit der Untersuchung wäre somit eine Überschätzung von Unterschieden zu erwarten.

c) Die gleiche statistische Auswertung wurde mit der unabhängigen Variable „Raum“ (Raum A/Raum B) durchgeführt, wobei es das Ziel war, eventuell auftretende Einflüsse der unterschiedlichen Untersuchungsumgebung sichtbar zu machen. Auch bei dieser statistischen Auswertung wurden lediglich die 86 Ergebnisse aus Testdurchgang 1 berücksichtigt.

Wenn die Grundvoraussetzungen für die Durchführung von χ^2 -Tests nicht erfüllt waren, wurden **Exakte Tests nach Fisher** durchgeführt.

Diese wurden verwendet,

a) um die Häufigkeitsverteilung von Übereinstimmungen und Nicht-Übereinstimmungen in Abhängigkeit von der Untersuchungsreihenfolge zu untersuchen. Dafür wurden einerseits jene Ergebnispaaire berücksichtigt, bei denen anfänglich Tester 1 und in der darauffolgenden zweiten Untersuchung Tester 2 untersuchte. Andererseits wurden jene Ergebnispaaire, die aus der umgekehrten Reihenfolge der Tester stammen, verwendet. Datengrundlage stellten die Untersuchungen 1 und 2 (21 Ergebnisvergleiche für die Reihenfolge A-B, bzw. 22 für die Reihenfolge B-A) dar. Unterschiede in der Anzahl der Übereinstimmungen könnten einen Hinweis auf einen möglichen Behandlungseffekt durch einen der beiden Tester geben.

b) Exakte Tests nach Fisher wurden ebenfalls verwendet, um die Häufigkeitsverteilungen der von den beiden Testern als sicher bewerteten Richtungen zu vergleichen (abhängige Variable: Untersuchungsergebnis, unabhängige Variable: Tester).

4. Ergebnisse

4.1 Interrater Reliabilität

Beim 1. Testdurchgang an 43 Probanden weicht der aus den Ergebnissen der beiden Osteopathen berechnete Kappa-Index für die Interrater Reliabilität signifikant von Null ab ($z=3,39$, $p=0,00070$), jedoch ist ein negativer Wert zu beobachten ($\kappa= -0,35$). Der Unterschied zwischen den Ergebnissen der beiden Tester ist somit signifikant höher, als durch deren zufällige Übereinstimmung erklärbar wäre. Dieses Ergebnis gibt folglich einen Hinweis auf systematische Unterschiede zwischen den beiden Untersuchern. Bei erwarteten fünfzehn Übereinstimmungen wurden nur fünf beobachtet. Die Einzelergebnisse der beiden Tester sind in "Interreliabilität im Testdurchgang 1" im Anhang zusammengefasst, die Anzahl beobachteter und erwarteter Häufigkeiten in Tab.3:

Tab.3: Beobachtete und erwartete Häufigkeit bei der Interrater Reliabilität

Beobachtete Übereinstimmung:

		Tester 2			Summe
		L	R	G	
Tester 1	L	2	15	5	22
	R	7	3	4	14
	G	4	3	0	7
		13	21	9	43

Erwartete Übereinstimmung:

		Tester 2			Summe
		L	R	G	
Tester 1	L	7	11	5	22
	R	4	7	3	14
	G	2	3	1	7
		13	21	9	43

Beobachtete Übereinstimmung		Erwartete Übereinstimmung	
N	P	n	P
5	0,12	15	5

Kappa: -0,35

Auch bei Betrachtung jener sechzehn Ergebnisse, die von zumindest einem der beiden Osteopathen als „sicher“ eingeschätzt wurden (siehe Tab.4), zeigt sich, dass nur bei einem einzigen Probanden eine Übereinstimmung zwischen den beiden Testern besteht. Der aus diesen sechzehn Datensätzen berechnete Kappa-Index beträgt $\kappa=-0,24$ und unterscheidet sich ebenfalls signifikant von Null ($z=-2,24$, $p=0,025$). Die beiden Tester stimmen also auch bei „sicheren“ Testergebnissen signifikant seltener überein, als durch eine zufällige Übereinstimmung zu erwarten wäre.

Tab.4: Beobachtbare und erwartete Häufigkeit bei der Interrater Reliabilität der sicheren Ergebnisse

Beobachtete Übereinstimmung:

		Tester 2			Summe
		L	R	G	
Tester 1	L	1	8	3	12
	R	2	0	0	2
	G	0	2	0	2
		3	10	3	16

Erwartete Übereinstimmung:

		Tester 2			Summe
		L	R	G	
Tester 1	L	2,3	7,5	2,3	12
	R	0,4	1,3	0,4	2
	G	0,4	1,3	0,4	2
		3	10	3	16

Beobachtete Übereinstimmung		Erwartete Übereinstimmung	
N	P	n	P
1	0,06	1	0,06

Kappa: -0,24

Selbst unter der Annahme dass die Antworten „gleich“ darauf zurückzuführen sind, dass das Ausmaß der Spannung nur gering ist und nur von einem der Tester richtig eingeschätzt, vom anderen jedoch nicht wahrgenommen wurde (d.h. nach Anpassung der Antworten „gleich“ an die Antwort des anderen Testers), ist bei Betrachtung aller 43 Ergebnisse nur eine zufällige Übereinstimmung zwischen den beiden Testern zu beobachten ($\kappa=0,011$, $z=0,0735$, $p=0,94$). Obwohl dieses Ergebnis keine positive Aussage über die Übereinstimmung der beiden Tester erlaubt, wird darauf hingewiesen, dass durch diese Vorgehensweise eine bessere Übereinstimmung zwischen den Testern erzwungen wurde.

Nachdem der **Testdurchgang 2** ebenfalls von beiden Osteopathen an den selben Probanden durchgeführt wurde, kann auch aus den hieraus gewonnenen Ergebnissen die Interrater-Reliabilität bestimmt werden. Cohen's Kappa beträgt bei Untersuchung 4 $\kappa=0,0042$, was zwar formal auf eine höhere Übereinstimmung hinweist, jedoch durch eine rein zufällige Übereinstimmung erklärbar ist ($z=0,0259$, $p=0,98$). Bei Untersuchung 3 ist hingegen mit $\kappa=-0,32$ eine sich signifikant von der zufälligen Übereinstimmung abhebende mangelnde Übereinstimmung zu beobachten ($z=-2,4$, $p=0,016$), die in derselben Größenordnung wie jene aus Testdurchgang 1 ermittelte liegt.

4.2 Test-Retest Reliabilität

Die Einzelergebnisse der beiden Tester bei den jeweils zwei Untersuchungen des zweiten Testdurchgangs sind im Anhang "Test Retest Reliabilität im Testdurchgang 2" zusammengefasst.

Bei **Tester 1** ist bei den Ergebnissen an jenen zwanzig Probanden, die im zweiten Testdurchgang (Untersuchungen 3 und 4) untersucht wurden, eine signifikante Abweichung ($z=3,31$, $p=0,00095$) des Kappa-Index von Null ($\kappa=0,57$) zu beobachten (siehe Tab.5). Bei erwarteten acht Übereinstimmungen wurden fünfzehn beobachtet. Das bedeutet, dass die Ergebnisse zu denen Tester 1 gelangte, an drei Viertel der Probanden/Probandinnen bei beiden Untersuchungen übereinstimmten. Nach Landis und Koch ist die Übereinstimmung als „mittelmäßig“ zu bezeichnen (Landis und Koch 1977).

Tab.5: Beobachtete und erwartete Häufigkeit der Test-Retest Reliabilität beim Tester 1

Beobachtete Übereinstimmung:

		Unters. 4			Summe
		L	R	G	
Unters. 3	L	9	2	0	11
	R	0	5	1	6
	G	2	0	1	3
		11	7	2	20

Erwartete Übereinstimmung:

		Unters. 4			Summe
		L	R	G	
Unters. 3	L	6	4	1	11
	R	3	2	1	6
	G	2	1	0	3
		11	7	2	20

Beobachtete Übereinstimmung		Erwartete Übereinstimmung	
N	P	n	P
15	0,75	8	0,42

Kappa: 0,57

Bei **Tester 2** ist im Gegensatz dazu keine signifikante Abweichung des Kappa-Index von Null zu beobachten ($z=-0,115$, $p=0,91$), der Kappa-Index beträgt $\kappa= -0,02$ und entspricht somit dem durch zufällige Übereinstimmung zu erwartenden Wert (siehe Tab.6). Es wurden neun Übereinstimmungen beobachtet, auf Basis von Zufall erwartet werden ebenfalls neun.

Trotz dieser Unterschiede besteht laut einem χ^2 -Test kein signifikanter Unterschied zwischen den bei den beiden Testern zu beobachtenden Häufigkeitsverteilungen von Übereinstimmungen und Nicht-Übereinstimmungen ($\chi^2=2,6042$, $df=1$, $p=0,11$).

Zur Einschätzung dieses Ergebnisses: Ein signifikantes Ergebnis wäre bei der vorliegenden Stichprobengröße mit einer um ein bis zwei Übereinstimmungen höheren Differenz zwischen

Tester 1 und 2 zu erwarten (z.B. Tester 1: sechzehn Übereinstimmungen/Tester 2: gleichbleibend neun Übereinstimmungen oder Tester 1: gleichbleibend fünfzehn Übereinstimmungen/Tester 2: sieben Übereinstimmungen).

Tab.6: Beobachtete und erwartete Häufigkeit der Test-Rest Reliabilität beim Tester 2

Beobachtete Übereinstimmung:

		Unters. 4			Summe
		L	R	G	
Unters. 3	L	1	3	0	4
	R	3	8	1	12
	G	2	2	0	4
		6	13	1	20

Erwartete Übereinstimmung:

		Unters. 4			Summe
		L	R	G	
Unters. 3	L	1	3	0	4
	R	4	8	1	12
	G	1	3	0	4
		6	13	1	20

Beobachtete Übereinstimmung		Erwartete Übereinstimmung	
N	P	n	P
9	0,45	9	0,46

Kappa: -0,02

4.3 Schmerzangaben und Untersuchungsergebnisse im Vergleich

Aus den bisher präsentierten Daten geht nicht hervor, ob einer der beiden Tester zu „richtigen“ Ergebnissen gekommen ist. Mangels einer vergleichbaren Messmethode wurden daher die Ergebnisse von den 21 Probanden, die eine Schmerzsymptomatik aufweisen, als Vergleichsdaten herangezogen (Siehe Anhang "Probandentabelle").

Die schmerzende Körperseite wird, wie von Snider (2008) untersucht, durch Körperspannung als weniger beweglich eingeschätzt. Der angewandte Test soll die besser bewegliche Seite identifizieren, so wurden jeweils die der schmerzenden Körperseite gegenüberliegenden Körperseiten den Testergebnissen (Testdurchgang 1) der einzelnen Osteopathen gegenübergestellt. Aus diesen Wertepaaren wurden die Kappa-Indices berechnet. Es ist allerdings darauf hinzuweisen, dass eine Spannungslokalisierung nicht notwendigerweise mit der Schmerzlokalisierung übereinstimmen muss.

Bei **Tester 1** beträgt der Kappa-Index $\kappa=0,11$ ($z=0,796$, $p=0,43$) (siehe Tab.7), bei **Tester 2** $\kappa=-0,16$ ($z=-1,01$, $p=0,31$) (siehe Tab.8). Das heißt, dass bei beiden Untersuchern die Übereinstimmung der Untersuchungsergebnisse mit den aufgrund der Schmerzlokalisierung

erwarteten Körperseiten, nicht signifikant über jene hinausgeht, die auf Basis von Zufall zu erwarten wäre.

Tab.7: Übereinstimmung der Schmerzangabe des Patienten mit dem Ergebnis von Tester 1

Beobachtete Übereinstimmung:

		Schmerzangabe des Pat.			Summe
		L	R	G	
Tester 1	L	3	4	4	11
	R	2	3	2	7
	G	0	1	2	3
		5	8	8	21

Erwartete Übereinstimmung:

		Schmerzangabe des Pat.			Summe
		L	R	G	
Tester 1	L	2,6	4,2	4,2	11
	R	1,7	2,7	2,7	7
	G	0,7	1,1	1,1	3
		5	8	8	21

Beobachtete Übereinstimmung		Erwartete Übereinstimmung	
N	P	n	P
8	0,38	6	0,31

Kappa: 0,11

Tab.8: Übereinstimmung der Schmerzangabe des Patienten mit dem Ergebnis von Tester 2

Beobachtete Übereinstimmung:

		Schmerzangabe des Pat.			Summe
		L	R	G	
Tester 2	L	1	2	3	6
	R	4	2	3	9
	G	0	4	2	6
		5	8	8	21

Erwartete Übereinstimmung:

		Schmerzangabe des Pat.			Summe
		L	R	G	
Tester 2	L	1,4	2,3	2,3	6
	R	2,1	3,4	3,4	9
	G	1,4	2,3	2,3	6
		5	8	8	21

Beobachtete Übereinstimmung		Erwartete Übereinstimmung	
N	P	n	P
5	0,24	7	0,34

Kappa: -0,16

Betrachtet man nur die Ergebnisse jener dreizehn Probanden, die eine eindeutige Körperseite angaben, auf der sie Schmerzen haben, (d.h. durch Elimination jener, die

beidseits Schmerzen aufweisen), beträgt der Kappa-Index bei Tester 1 $\kappa=0,032$ ($z=-0,139$, $p=0,89$) (siehe Tab.9) und bei Tester 2 $\kappa=-0,23$ ($z=-1,32$, $p=0,19$) (siehe Tab.10).

Tab.9: Übereinstimmung der Rechts oder Links Schmerzangabe des Patienten mit dem Ergebnis von Tester 1

Beobachtete Übereinstimmung:

		Schmerzangabe des Pat.			Summe
		L	R	G	
Tester 1	L	3	4	0	7
	R	2	3	0	5
	G	0	1	0	1
		5	8	0	13

Erwartete Übereinstimmung:

		Schmerzangabe des Pat.			Summe
		L	R	G	
Tester 1	L	2,7	4,3	0,0	7
	R	1,9	3,1	0,0	5
	G	0,4	0,6	0,0	1
		5	8	0	13

Beobachtete Übereinstimmung		Erwartete Übereinstimmung	
N	P	n	P
6	0,46	6	0,44

Kappa: 0,032

Tab.10: Übereinstimmung der Rechts oder Links Schmerzangabe des Patienten mit dem Ergebnis von Tester 2

Beobachtete Übereinstimmung:

		Schmerzangabe des Pat.			Summe
		L	R	G	
Tester 2	L	1	2	0	3
	R	4	2	0	6
	G	0	4	0	4
		5	8	0	13

Erwartete Übereinstimmung:

		Schmerzangabe des Pat.			Summe
		L	R	G	
Tester 2	L	1,2	1,8	0,0	3
	R	2,3	3,7	0,0	6
	G	1,5	2,5	0,0	4
		5	8	0	13

Beobachtete Übereinstimmung		Erwartete Übereinstimmung	
N	P	n	P
3	0,23	5	0,37

Kappa: -0,23

Bei beiden Testern ist also eine weitere Verschlechterung der Übereinstimmung zu beobachten. Es ist allerdings darauf hinzuweisen, dass die Antwortmöglichkeiten der Tester

weiterhin auch die Kategorie "gleich" beinhalten, was eine Verletzung der Grundbedingungen für die Berechnung des Kappa-Index darstellt. Bei der Berechnung des Kappa-Index muss sichergestellt sein, dass die Probanden alle möglichen Testkategorien aufweisen.

Eine Elimination der Ergebnisse „gleich“ aus den Testergebnissen würde allerdings ebenfalls eine Verfälschung darstellen, da die Tester ohne diese Alternative zu einem anderen Ergebnis hätten kommen müssen. In diesem Fall erhält man aus den von Tester 1 an zwölf Probanden getesteten Ergebnissen einen κ -Wert von $\kappa=0,030$ und von Tester 2 (neun Probanden) $\kappa= -0,29$, also eine übereinstimmende Aussage.

Aufgrund der geringen Anzahl an Ergebnissen und der oben diskutierten Einschränkungen können diese Ergebnisse jedoch nur als *Hinweis* darauf gewertet werden, dass selbst bei theoretisch eindeutigen Spannungsverhältnissen keine wesentliche Übereinstimmung besteht.

4.4 Ursachenforschung für die systematischen Unterschiede

4.4.1 Ergebnisse der Feedbackbögen der Probanden

Aus den Feedbackbögen der Probanden (siehe Anhang) ist ersichtlich, dass 32 der 40 Probanden, die diese ausfüllten, **Unterschiede in der Griffhaltung** der beiden Untersucher wahrgenommen haben. Vier Probanden gaben an, dass es sich um geringe Unterschiede handelte. Jeweils einer gab an, dass ein Tester die Finger weiter auseinander hatte, als der andere, dass ein Tester mehr mittig, der andere mehr seitlich getestet hat und dass einer mehr „von unten rauf“, der andere „mehr auf Rippe“ gedrückt habe.

Auch hinsichtlich der **Druckstärke** sind 30 der 40 Probanden, die die Feedbackbögen ausfüllten der Ansicht, dass es Unterschiede zwischen den beiden Untersuchern gegeben habe, wobei zwei angaben, dass diese nur minimal waren. Vier gaben an, dass einer fester gedrückt habe, zwei, dass eine Person deutlich kräftiger war, und drei schließen aus den Unterschieden fälschlicherweise, dass ein Untersucher ein Mann und der andere eine Frau gewesen sei.

Eine weitere Aussage zeigt, dass jener Untersucher, der „deutlich sanfter“ war, auf die Atmung gewartet hat und beide Körperseiten untersucht hat, während der andere einen stärkeren Druck ausübte und (zumindest bei diesem Probanden) die Untersuchung nur auf eine Seite beschränkte.

Auch bereits bei der Einschulung der Osteopathen waren zu Beginn Stärke und Art der Griffhaltung unterschiedlich. Anhand des Feedbacks von der Autorin und eines Probanden wurde eine Angleichung erreicht. Es liegt die Vermutung nahe, dass durch die Routine die eingelernte von der automatisierten Untersuchungstechnik abgelöst wurde.

Insgesamt hatten nur fünf der 40 Probanden, die die Feedbackbögen ausfüllten, das Gefühl vom **gleichen Osteopathen** untersucht worden zu sein.

Das **Befinden während der und nach den Tests** wurde von allen Probanden als Wohlfühl angegeben, Schmerz wurde von niemandem empfunden.

Im Gespräch mit den Untersuchern stellte sich heraus, dass die Mehrheit der Probanden während der Untersuchung nicht entspannt, sondern angespannt, teilweise mit angehaltenem Atem oder maximaler Ein- und Ausatmung gelegen ist.

4.4.2 Statistische Untersuchung eines möglichen Behandlungseffekts

Anhand des Probandenfeedbacks ist ersichtlich, dass ein Untersucher, der anhand der konkreten Aussagen und der Untersuchungsreihenfolge als Tester 2 identifiziert werden konnte, einen stärkeren Druck ausübte (und sich zumindest bei ein paar Probanden auch in anderen Details von der Testdurchführung des Testers 1 unterschied). Die Daten wurden daher hinsichtlich eines möglichen Behandlungseffekts untersucht, der manchmal zur Rechtfertigung in der Literatur wie z.B. bei Bengaard et al (2012) von ungenügenden Übereinstimmungen verwendet wird.

Dazu floss die Reihenfolge, in der die beiden Tester die Probanden untersuchten, in die Auswertung mit ein. Die Kappa-Indizes wurden getrennt berechnet für die Probanden, die in aufeinanderfolgenden Untersuchungen zuvor von Tester 1 und anschließend von Tester 2 untersucht wurden, und für jene, die zuvor von Tester 2 und anschließend von Tester 1 untersucht wurden. Die Einzelergebnisse der beiden Tester sind gegliedert nach der Untersuchungsreihenfolge der Tester im Anhang zusammengefasst, die Anzahl beobachteter und erwarteter Häufigkeiten in Tab.11 und Tab.12.

Tab. 11: Beobachtete und erwartete Übereinstimmung der Untersuchungsreihenfolge Tester 2 - Tester 1 (Interrater Reliabilität)

Beobachtete Übereinstimmung:

		Tester 1			Summe
		L	R	G	
f _{Tester 2}	L	0	4	1	5
	R	8	1	3	12
	G	3	1	0	4
		11	6	4	21

Erwartete Übereinstimmung:

		Tester 1			Summe
		L	R	G	
Tester 2	L	2,6	1,4	1,0	5
	R	6,3	3,4	2,3	12
	G	2,1	1,1	0,8	4
		11	6	4	21

Beobachtete Übereinstimmung		Erwartete Übereinstimmung	
N	P	n	P
1	0,05	7	0,32

Kappa: -0,41

Tab. 12: Beobachtete und erwartete Übereinstimmung der Untersuchungsreihenfolge Tester 1 - Tester 2 (Interrater Reliabilität)

Beobachtete Übereinstimmung:

		Tester 2			Summe
		L	R	G	
Tester 1	L	2	7	2	11
	R	3	2	3	8
	G	3	0	0	3
		8	9	5	22

Erwartete Übereinstimmung:

		Tester 2			Summe
		L	R	G	
Tester 1	L	4,0	4,5	2,5	11
	R	2,9	3,3	1,8	8
	G	1,1	1,2	0,7	3
		8	9	5	22

Beobachtete Übereinstimmung		Erwartete Übereinstimmung	
N	p	n	P
4	0,18	8	0,36

Kappa: -0,28

Der aus den Daten jener 21 Probanden, die im ersten Testdurchgang bei der Untersuchung 1 für die Bestimmung der Interrater-Reliabilität von Tester 2 und im bei der Untersuchung 2 von Tester 1 untersucht wurden, berechnete Kappa-Index für die Interrater-Reliabilität

beträgt $\kappa=-0,41$ (siehe Tab 11) und unterscheidet sich signifikant von Null ($z=-2,98$, $p=0,0029$).

Auch bei umgekehrter Untersuchungsreihenfolge, d.h. bei anfänglicher Untersuchung von Tester 1 und anschließender von Tester 2 an 22 Probanden, ist ein negativer Kappa-Index für die Interrater-Reliabilität zu beobachten ($\kappa=-0,28$), der sich auf dem gewählten Signifikanzniveau von $\alpha=0,05$ nicht signifikant, jedoch tendenziell von Null unterscheidet ($z=-1,84$, $p=0,065$) (siehe Tab.12).

Das heißt, dass sich die Übereinstimmung unabhängig von der Untersuchungsreihenfolge zumindest tendenziell von einer auf Zufall beruhenden Übereinstimmung unterscheidet.

Auch in der Häufigkeitsverteilung von Übereinstimmungen und Nicht-Übereinstimmungen ist keine signifikante Abhängigkeit von der Untersuchungsreihenfolge erkennbar (Exakter Test nach Fisher: $p=0,34$). Das deutet darauf hin, dass ein etwaiger Behandlungseffekt entweder durch beide Untersucher in ähnlicher Größenordnung, oder bei keinem der beiden auftritt.

4.4.3 Statistische Untersuchung von Einflüssen der Räumlichkeiten

Nachdem die Untersuchungen parallel in zwei Räumen (Raum A und Raum B) durchgeführt wurden, besteht theoretisch die Möglichkeit, dass beispielsweise Unterschiede im Raumangebot für die Tester oder andere Unterschiede in der Umgebung einen Einfluss auf die Ergebnisse haben könnten. Die Ergebnisse der beiden Tester sind, gegliedert nach dem Raum, in dem sie gewonnen wurden, im Anhang "Mögliche Einflüsse der unterschiedlichen Untersuchungsräume" zusammengefasst.

Auch wenn diese Möglichkeit dadurch, dass beide Tester sowohl in Raum A, als auch Raum B Untersuchungen durchführten, bei der Studiendurchführung Berücksichtigung fand, wurde untersucht, ob sich die Häufigkeit der in den beiden Räumen ermittelten Ergebnisse voneinander unterscheiden.

Ein χ^2 -Test der Häufigkeiten der in den beiden Räumen getesteten Ergebnisse "links", "rechts" und "gleich" (siehe Tab.13) ergab $\chi^2=0,2393$, $df=2$, $p=0,89$. Es gibt somit keinen Hinweis auf einen Einfluss der Untersuchungsumgebung auf die getestete Spannungsseite vom Diaphragma. Bei diesem statistischen Test wurden nur die Ergebnisse des ersten Testdurchgangs berücksichtigt.

Tab.13: Gegenüberstellung der Häufigkeit der Untersuchungsergebnisse in den beiden Räumen von den beiden Testern

	Raum A			Raum B		
	Tester 1	Tester 2	Summe	Tester 1	Tester 2	Summe
Links	11	8	19	11	8	16
Rechts	8	9	17	8	9	18
Gleich	3	5	8	3	5	8

4.4.4 Häufigkeit der Ergebnisse in Abhängigkeit vom Tester

Einen weiteren möglichen Grund für unterschiedliche Untersuchungsergebnisse stellen (unterbewusste, oder möglicherweise auch durch eigene Körperspannungen ausgelöste) Präferenzen der Tester dar. Leider wurde zu diesem Thema trotz intensiver Recherche in der Literatur keine Aussage gefunden. Für die Untersuchung möglicher Präferenzen wurde die Häufigkeit, mit der die einzelnen Spannungsseiten des Diaphragmas angegeben wurden, in Abhängigkeit vom Tester analysiert (siehe Tab.14). Für diese statistische Untersuchung wurden die Ergebnisse jedes Testers aus der Untersuchung 1 bzw. 2 (Testdurchgang 1) verwendet.

Tab.14: Gegenüberstellung der Häufigkeit der Untersuchungsergebnisse der beiden Tester im Testdurchgang 1

	Tester 1	Tester 2
Links	22	13
Rechts	14	21
Gleich	7	9

Während Tester 1 vorwiegend (22x) Spannungen in der linken Körperhälfte und vierzehn in der rechten vorfand, fand Tester 2 am häufigsten Spannungen in der rechten Körperhälfte vor (21x), während er in der linken auf nur dreizehn kam.

Es wurden von der Autorin keine Studien oder Literatur gefunden, ob die Körperspannung rechts und links bei gesunden Menschen gleich ist. Allerdings wurde im Unterricht von Christian Fossum ebenso wie bei Dr. Erich Mayer-Fally darauf aufmerksam gemacht, dass aufgrund der Untersuchungserfahrung rechts mehr Spannung zu erwarten ist. Als Hypothese wurde die Fülle und das erhöhte Eigengewicht der Leber im Vergleich zum Magen angeführt. Anhand der Ergebnisse eines χ^2 -Tests sind diese Unterschiede in der Häufigkeitsverteilung der möglichen Untersuchungsergebnisse („links“, „rechts“, „gleich“) der beiden Tester jedoch nicht signifikant ($\chi^2=3,9643$, $df=2$, $p=0,14$).

Sieht man sich an, welche Ergebnisse von den beiden Testern als „sicher“ bezeichnet wurden (siehe Tab.15), sind das bei Tester 1 ebenfalls vorwiegend Spannungen in der linken Körperhälfte (neun der elf von diesem Tester als sicher eingeschätzten Untersuchungsergebnisse), bei Tester 2 in der rechten (fünf der sechs von diesem Tester als sicher eingeschätzten Untersuchungsergebnisse). Die Originaldaten sind im Anhang "Häufigkeit der Untersuchungsergebnisse in Abhängigkeit der Tester" ersichtlich.

Tab.15: Gegenüberstellung der Häufigkeit der von den beiden Testern als „sicheres“ Ergebnis bezeichneten Untersuchungsergebnisse

	Tester 1	Tester 2
Links	9	1
Rechts	2	5
Gleich	0	0

In einem Exakten Test nach Fisher stellt sich der zwischen den Testern bestehende Unterschied in der Häufigkeit der als „sicheres“ Untersuchungsergebnis bewerteten getesteten Spannungsausrichtungen des Diaphragmas („Links“, „Rechts“ und „Gleich“) als signifikant heraus ($p=0,035$).

4.5 Überprüfung der Hypothesen

4.5.1 Hypothese 1

"Führen zwei Osteopathen innerhalb von zwei Stunden am selben Probanden einen Test am abdominalen Diaphragma in Translation aus, ist die Übereinstimmung ihrer Ergebnisse statistisch signifikant."

Die Interrater Reliabilität ist lediglich ($\kappa=-0,35$), das heißt Untersucher kommen seltener zu den selben Ergebnissen, als durch zufällige Übereinstimmung zu erwarten wäre. Der Unterschied ist jedoch signifikant ($p=0,0007$). Es gibt einen Hinweis darauf, dass die mangelnde Übereinstimmung auf systematischen Unterschieden zwischen den Testern beruht. Das gilt auch für die als „sicher“ bewerteten Testergebnisse ($\kappa=-0,24$, $p=0,025$). Hat zumindest ein Osteopath sein Ergebnis als "sicher" eingeschätzt, liegt bei vorhandenen sechzehn Datensätzen ein Kappaindex von $\kappa=-0,24$ vor. Selbst wenn man die Antwort „gleich“ dem Ergebnis des jeweils anderen Testers anpasst, und somit eine Verbesserung der Übereinstimmung erzwingt (was in der Diskussion betrachtet werden wird), kann nur ein Ergebnis, das jenem zufälliger Übereinstimmung entspräche ($\kappa=0,011$, $p=0,94$) beobachtet werden.

Daher muss die aufgestellte Hypothese 1 abgelehnt werden.

4.5.2 Hypothese 2

"Führt ein Osteopath innerhalb von zwei Stunden am selben Probanden einen Test am abdominalen Diaphragma in Translation aus, ist die Übereinstimmung seiner Ergebnisse statistisch signifikant."

Bei der Test-Retest Reliabilität erreicht Tester 1 bei 75% Übereinstimmungen ($\kappa=0,57$, $p=0,00095$), Das ergibt laut Landis und Koch eine mittelmäßige Übereinstimmung. Tester 2 hat eine 45%ige Übereinstimmung in seinen Ergebnissen ($\kappa=-0,02$, $p=0,91$). Das entspricht einer schwachen Übereinstimmung laut Landis und Koch. Ein signifikantes Ergebnis wäre bei der vorliegenden Stichprobengröße mit einer um ein bis zwei Übereinstimmungen höheren Differenz zwischen Tester 1 und 2 zu erwarten. Die Tester unterscheiden sich

voneinander statistisch nicht signifikant ($p=0,11$) in der Häufigkeit von Übereinstimmungen und Nicht-Übereinstimmungen.

Für Tester 1 kann Hypothese 2 als angenommen werden, für Tester 2 nicht.

4.5.3 Hypothese 3

"Die Übereinstimmung der Ergebnisse bei Patienten mit chronischen Schmerzen im lumbalen, thorakolumbalen oder thorakalen Bereich ist im Vergleich zu schmerzfreien Probanden statistisch signifikant."

Bei beiden Untersuchern geht die Übereinstimmung der Untersuchungsergebnisse mit der aufgrund der angegebenen Schmerzseite des Probanden erwarteten Körperseite nicht über den Zufall hinaus (Tester 1 mit $\kappa=0,11$, $p=0,43$, Tester 2 mit $\kappa=-0,16$, $p=0,31$). Konkret heißt das bei Tester 1 acht beobachtete von sechs erwarteten Übereinstimmungen und bei Tester 2 fünf beobachtete von sieben erwarteten Übereinstimmungen.

Hypothese 3 kann somit nicht bestätigt werden.

5. Diskussion

5.1 Diskussion zur Hypothese 1

Die beiden Tester, die sogar seltener als durch Zufall in ihren Ergebnissen übereinstimmen, geben einen signifikanten Hinweis auf systematische Unterschiede. Für die systematischen Unterschiede zwischen den Testern bezüglich der Interrater Reliabilität wurden die Ergebnisse der Feedbackbögen der Probanden ausgewertet und eine statistische Analyse in Bezug zu einem möglichen Behandlungseffekt durchgeführt. Weiters wurden mögliche Einflüsse der Untersuchungsräumlichkeiten statistisch abgeklärt und die Häufigkeit der Testergebnisse (Körperseiten) von Tester 1 und Tester 2 verglichen.

1. Die Mehrheit der Probanden gibt in den Feedbackbögen Unterschiede in der Griffhaltung (32 von 40) und der Druckstärke (30 von 40) an. Nur fünf von 40 Probanden glaubten vom gleichen Osteopathen untersucht worden zu sein. Der in den Feedbackbögen beschriebene Unterschied zwischen den beiden Testern könnte Auswirkungen auf die unterschiedlichen Interratertestergebnisse haben, ist aber statistisch nicht signifikant.
2. Unabhängig davon, ob Tester 1 oder Tester 2 als erster testete, ist eine **negative** signifikante (Tester 2 mit $\kappa=-0,41$, $p=0,0029$) oder tendenzielle (Tester 1 mit $\kappa=-0,28$, $p=0,065$) Abweichung von $\kappa=0$ (zufällige Übereinstimmung) zu beobachten. Auch in der Häufigkeitsverteilung von Übereinstimmungen und Nicht-Übereinstimmungen ist keine signifikante Abhängigkeit von der Untersuchungsreihenfolge erkennbar (Exakter Test nach Fisher: $p=0,34$). Ein Behandlungseffekt ist daher entweder nicht vorhanden, oder tritt bei beiden Testern in ähnlicher Größenordnung auf.
3. Ausgeschlossen werden kann ein Einfluss unterschiedlicher Umgebungseinflüsse in den beiden Behandlungsräumen. Diese Möglichkeit wurde bereits bei der Testdurchführung dadurch berücksichtigt, dass beide Tester sowohl in Raum A, als auch Raum B Untersuchungen durchführten. Ein χ^2 -Test der Häufigkeiten der in den beiden Räumen getesteten Ergebnisse „links“, „rechts“ und „gleich“ ergab $\chi^2=0,2393$, $df=2$, $p=0,89$ und bestätigte somit, dass kein Einfluss der Untersuchungsumgebung auf die getestete Spannungsseite vom Diaphragma besteht.
4. Individuelle Präferenzen können aus den als „sicher“ bezeichneten Untersuchungsergebnissen abgeleitet werden. Hier gibt Tester 1 neunmal „links“ und nur zweimal „rechts“ an, während Tester 2 fünfmal das Ergebnis „rechts“ und einmal das Ergebnis „links“ als „sicher“ anführt. Als signifikant ($p=0,035$) stellt sich nach

einem Exakten Test nach Fisher der Unterschied beider Tester in der Häufigkeit der als sicher bewerteten getesteten Spannungsausrichtungen des Diaphragmas heraus. Die beiden Tester stimmten nur bei einem Probanden darin überein, ein sicheres Ergebnis gefunden zu haben. Auf Grund der statistischen Analyse ist anzunehmen, dass Tester 1 bevorzugt Spannungen in der linken Körperhälfte und Tester 2 in der rechten wahrnimmt.

Diese möglicherweise präferierten Körperrichtungen spiegeln sich auch in den gesamten Untersuchungsergebnissen des Testdurchgangs 1 wieder. Während Tester 1 vorwiegend (22x) Spannungen in der linken Körperhälfte und vierzehn in der rechten vorfand, fand Tester 2 am häufigsten Spannungen in der rechten Körperhälfte vor (21x), während er in der linken auf nur dreizehn kam. Das Ergebnis „Gleich“ wurde von Tester 1 siebenmal und von Tester 2 neunmal vorgefunden. Allerdings sind diese Unterschiede zwischen den beiden Testern in der Häufigkeitsverteilung der möglichen Untersuchungsergebnisse („links“, „rechts“, „gleich“) nicht signifikant ($\chi^2=3,9643$, $df=2$, $p=0,14$).

Nach Oschman (2002) können unterschiedliche Untersucher unterschiedliche Symptome vom gleichen Probanden erfassen und daraufhin unterschiedlich interpretieren. Jeder Mensch hat laut Oschman seine eigenen biomagnetischen Felder, die bei Palpation verschieden interferieren können. Es macht also einen Unterschied, ob Untersucher 1 oder Untersucher 2 mit demselben Probanden in Kontakt tritt, da je nach Leitung andere Felder angesprochen werden. So ist ein anderes Untersuchungsergebnis, in weiterer Folge eine andere Gewichtung in der Diagnose und ein anderer Behandlungsansatz begründbar.

Sommerfeld et al (2004) setzt in seiner Studie ebenfalls die Präsenz des Untersuchers in den Vordergrund, da der Atemrhythmus vom Untersucher mit dem palpieren cranialen Rhythmus vom Probanden korreliert. Dies würde, umgemünzt auf die vorliegende Studie, zwar eine schlechte Interrater Reliabilität begründen, müsste jedoch eine bessere Test-Rest Reliabilität mit sich ziehen.

Ebenso zeigt Norton (1996) eine signifikante Wechselseitigkeit zwischen dem cranialen und sacralen Rhythmus beim selben Tester (es wurde bei dieser Studie am Probanden beim Cranium und beim Sacrum jeweils von einem Tester der craniale Rhythmus untersucht). Vielleicht beeinflusste auch bei der vorliegenden Studie der eigene Rhythmus des Testers das Ergebnis und führte so zu unterschiedlichen Deutungen.

5.2 Diskussion zur Hypothese 2

Tester 1 weist eine höhere signifikante Reproduzierbarkeit auf ($\kappa=0,57$, $p=0,00095$), jedoch ist weder eine Übereinstimmung mit dem anderen Tester gegeben (Interrater Reliabilität $\kappa=0,35$, $p=0,0007$), noch eine Übereinstimmung mit der erwarteten Körperseite von symptomatischen Patienten ($\kappa=0,11$, $p=0,31$). Theoretisch erreicht ein Tester, der immer dieselbe bevorzugte Körperseite angibt, die höchste Reproduzierbarkeit, auch wenn die Ergebnisse objektiv falsch sind. Es ist daraus also nicht abzuleiten, dass Tester 1 richtig und Tester 2 falsch getestet hat. Bei mehreren Testern kann man zumindest theoretisch anhand der Mehrheitsergebnisse abschätzen, wer nicht dazu passt.

Aufgrund der Ergebnisse dieser Studie ist es empfehlenswert, Reliabilitätsstudien mit mehr als zwei Testern durchzuführen - was auch der ursprüngliche Gedanke zur vorliegenden Masterthese war. Die Reproduzierbarkeit von Tester 1 ist hoch, die von Tester 2 niedrig. Wer von den beiden nun für die Gesamtheit der Osteopathen charakteristisch ist, kann mit dieser Studie nicht beantwortet werden. So kann nur vermutet werden, dass es Osteopathen gibt, für die dieser Test reproduzierbar ist, für andere nicht. Eine höhere Untersucherzahl würde statistische Ausreißer sichtbar machen.

Die persönliche Vermutung der Autorin zur schlechteren Kongruenz der Ergebnisse von Tester 2 liegt in der Konzentration auf die Testausführung. Vielleicht wurde in der einmonatigen Einübungsphase der Test anders eingelernt, was zwar in der Einschulung vor der Untersuchung angeglichen werden konnte, aber durch die darauffolgende Konzentration auf die Ausführung von der Palpationsfähigkeit ablenkte (siehe "Ausblick").

Hätten die Untersucher beim Ergebnis anstatt zwischen "gleich, rechts und links" nur zwischen "rechts und links" gewählt, wäre vielleicht bei Grenzfällen die Entscheidung leichter gefallen und hätte das intuitive Gefühl mehr die Oberhand behalten.

Prinzipiell kann die Vergleichbarkeit von der Spannung des rechten und linken Diaphragmas in Frage gestellt werden, da durch die unterschiedlichen darunterliegenden Organe mit der rechtsseitigen Leber und dem linksseitigen Magen andere Bedingungen herrschen. Die Fülle und Schwere der Leber kann vom Hohlorgan Magen von den Untersuchern unterschiedlich aufgefasst werden. Die Erfahrung von Dr. Erich Mayer-Fally und Christian Fossum bewertet bei einem gesunden Menschen die rechte Diaphragmenseite mit mehr Spannung. Diese Erfahrung korreliert mit dem Muster von Zink, das zu 80% im Beckendiaphragma rechts, im abdominalen Diaphragma links, beim Thoracic Outlet rechts und beim Tentorium Cerebelli links rotiert ist (Pope, 2003). Durch die vermehrte Spannung rechts nimmt das Diaphragma den leichteren Weg und dreht nach links.

5.3 Diskussion zur Hypothese 3

Als Prüfung der Richtigkeit der Palpation kann die Beziehung von Schmerzsymptomatik und Spannung gesehen werden. Snider et al (2008) und Kuchera (2007) sind nur zwei von vielen Referenzen, die den Schmerz zwar als Behandlungsgrund für den Patienten sehen, jedoch darauf hinweisen, dass der Schmerz nur ein Parameter neben Gewebsveränderung, Asymmetrie, Restriktion und Sensitivität für die Dysfunktion ist. Projektionen, Kompensationen oder psychische Komponenten können bei Alleinbetrachtung des Schmerzes und der Spannung übersehen werden. Der Autorin ist bewusst, dass die Gleichsetzung der Spannungslokalisation mit der Schmerzlokalisierung eine Reduktion darstellt.

Bengaard et al (2012) nimmt für die Frage, ob "richtig oder falsch" getestet wurde, zusätzlich zu zwei lebendigen Modellen ein Plastikmodell mit einer fixierten Seite mit in die Testung auf. Die Übereinstimmung in Bengaards Studie für die Palpation der Position von der Spina Iliaca anterior superior war zwischen den 151 Testern trotzdem schlecht ($\kappa=0,038$, $p=0,453$). Wegen dem hohen Aufwand an Vorbereitungen wurde die Plastikmodellidee schlussendlich nicht in die vorliegende Studie übernommen.

Zu überlegen ist die Eindeutigkeit der alleinigen subjektiven Schmerzangabe der Probanden ohne zusätzliche ärztliche Befunde. In der Diskussion mit den Probanden ergab sich teilweise die eigene Unsicherheit in der Zuordnung des chronischen Schmerzes auf eine bestimmte Seitenzugehörigkeit oder Beidseitigkeit.

Vielleicht hätte ein zusätzlicher Palpationstest in Bauchlage zur Feststellung der Gewebsveränderung der autochthonen Rückenmuskulatur die subjektive Schmerzangabe der Probanden bestätigen können. Eine Kombination von verschiedenen Tests macht die Untersuchung für Zusammenhänge offener, jedoch auch komplexer.

Obwohl ein Zusammenhang von Blockaden der Wirbelsäule zu Änderungen der Diaphragmaspannung nicht vorauszusetzen ist, wurde eine Voruntersuchung der Probanden auf Blockaden in der Wirbelsäule als Hinweis, ob Schmerz und Restriktion korrelieren, in Erwägung gezogen. Die Untersuchung im Sitzen durch die Autorin wäre allerdings wieder als subjektiver Eindruck zu werten gewesen, da der Test noch keine Validität vorweisen kann, und darum nicht durchgeführt.

Aufgrund der Unsicherheit, ob der Schmerz auf der Gewebsebene ipsilateral korreliert, oder doch wie oben beschrieben eine Kompensation, Projektion oder eine psychischen Part übernimmt, wäre es vielleicht besser gewesen, Schmerzpatienten auszuschließen. Möglicherweise hätte das Ausklammern des Parameters Schmerz, also die alleinige Untersuchung von symptomfreien Probanden, eine übersichtlichere Reliabilitätstestung abgegeben.

5.4 Kritik an der eigenen Studie

Nach Aussage der Untersucher konnten die Probanden die Anforderung, entspannt zu liegen und gleichmäßig zu atmen, nicht einhalten. Die erhöhte Grundspannung zum Beispiel durch das Anhalten der Atmung bzw. maximales Ein- und Ausatmen der Probanden erschwerte die Testung. Vielleicht hätte das angeleitete Ein- und Ausatmen durch die Assistenten und die Testung in Einatmung zum Vereinheitlichen der Ausgangsposition und zur Entspannung beitragen können. Durch die erhöhte Grundspannung des Probanden könnte der geringe translatorische Bewegungsimpuls vom Untersucher im Test zu sensitiv gewesen sein und so die Reproduzierbarkeit und Aussagekraft minimiert haben.

Ein Schwachpunkt der vorliegenden Studie besteht in der unvollständigen Blindierung der Untersucher. Leider wurde bei der Untersuchung verabsäumt, den Probanden beckenabwärts mit einem Leintuch zuzudecken, wie es im Probedurchlauf gemacht wurde. Die Rückfrage bei den Testern ergab jedoch, dass durch die hohe Konzentration beim Untersuchen wie durch die hohe Probandenzahl (es wurden in zwei Stunden von jedem Osteopathen 63 Untersuchungen durchgeführt) keine Chance auf Wiedererkennung der Probanden mit Erinnern des Untersuchungsergebnisses gegeben war.

Ein weiterer Aspekt soll zur Einschulung der Osteopathen erwähnt werden. Durch die mündliche und schriftliche Testanleitung einen Monat vor der Studie ist trotz genau definierter und beschriebener Testdurchführung (die Untersucher bekamen die Formulierung von Kapitel 3.1. zugesendet.) ein Rest an Eigeninterpretation der Untersucher in Erwägung zu ziehen. Es könnte eine individuelle Form der Testausführung im Probemonat stattgefunden haben. Durch die praktische Einschulung vor der Untersuchung könnte eine "Nuancenveränderung" der eingeübten Testausführung stattgefunden haben, die von der Beurteilung des Testergebnisses ablenkte, da die Aufmerksamkeit wieder vermehrt auf die Ausführung des Tests gelenkt wurde. Die Aussagekraft des Tests könnte so herabgesetzt worden sein.

Ebenfalls zu bedenken ist, vielleicht auch durch die unterschiedliche Testausführung, die Möglichkeit der unterschiedlichen Palpationstiefe der beiden Untersucher. Angesichts der Feedbackbögen der Probanden mit dem Ergebnis unterschiedlicher Druckstärke beider Tester könnten divergente Gewebe untersucht worden sein.

Nach Liem (2006, S.151) wird mit leichtem Druck eine Information über Tonus, Elastizität und Turgor von der Haut gewonnen. Etwas festerer Druck gibt Auskunft über Tonus und Turgor der oberflächlichen Muskelschichten und tieferer Druck informiert über tiefere Muskelschichten, fasziale Umhüllungen und den Organzustand. Im Fall dieser Studie könnte ein Untersucher den Spannungszustand der Intercostalmuskulatur, der andere den vom Diaphragma getestet haben.

Im Sinne der Tensegrity müsste bei unterschiedlicher Palpationstiefe trotzdem ein gleiches Ergebnis erzielt werden, da das System die Spannung auf alle Ebenen aufteilt. Ist die Intercostalmuskulatur links erhöht, müsste auch das linke Diaphragma vermehrt in Spannung sein (wenn man von der Weiterleitung in den Körperdiagonalen durch Kreuzung der Fasern der Körpermitte absieht, siehe Kapitel 2.3.1.).

Ein zusätzliches Manko der Studie könnte der groß gewählte Schmerzbereich vom Brust- bis Lendenwirbelsäulenbereich sein. Durch das Tensegritymodell könnte jedoch vom Schmerz der umliegenden Wirbelsäulenbereichen eine Weiterleitung zur Diaphragmaspannung stattfinden. Vielleicht hätte trotzdem eine Einengung der Symptomatik im Thorakolumbalen Übergang übereinstimmendere Ergebnisse geliefert, was aber das Erreichen der Probandenzahl deutlich erschwert hätte.

5.5 Ausblick

Eine Idee zur Erklärung individueller Präferenzen beider Untersucher wäre eine unterschiedliche Händigkeit. Beide Untersucher gaben jedoch an, bimanuell zu sein, mit der Festlegung rechts zu schreiben. Eine interessante Studie wäre in dem Fall, ob eine Untersuchung mit Rechts- und Linkshändern auch individuelle Präferenzen bei einem reliablen Test ergibt. In der Literatur konnte zu diesem Thema nichts gefunden werden.

Eine andere Möglichkeit wäre eine Untersuchung des Standorts des Testers während der Untersuchung. Ergeben sich auch Präferenzen abhängig von seiner Händigkeit, wenn der Untersucher den Test von der rechten beziehungsweise linken Seite neben dem Behandlungsbett ausführt?

Das Aufheben der Blindierung könnte die erhöhte Grundspannung der Probanden herabsetzen. Die Probanden wären vielleicht gefasster auf die Berührung und so eventuell entspannter im Grundtonus, was die Palpation erleichtern würde. Zudem könnten, wie von Krause (2008) beschrieben, die Untersucher nicht nur ihren Tastsinn, sondern alle Sinne für das Ergebnis nützen. Vielleicht wäre in diesem Fall eine alleinige Untersuchung der Interrater Reliabilität ausreichend. Eine Studie, die ähnliche Parameter wie die vorliegende Masterthese aufweist und damit vergleichbar ist, würde mit Aufheben der Blindierung über diese Mutmaßung Aufschluss geben.

Eine klinisch relevante Information für die Diagnose kann nur aus einer Kombination von mehreren Tests gefiltert werden. Patijn (2004) warnt allerdings bei Testkombinationen vor der Abhängigkeit der Ergebnisse zwischen den ausgewählten Tests. Snider et al (2008) konnte übereinstimmende Testergebnisse bei "Low back pain" Patienten von zwei Untersuchern erreichen, indem er vier verschiedenen Parametern beurteilte ("resistance to anterior springing, tenderness, tissue texture change, static rotational asymmetry"). Liem et al (2012) schlägt ein standardisiertes Diagnoseverfahren (ein Aneinanderreihen von

mehreren Tests) vor, um zehn Osteopathen zu vergleichen. Eine komplexe Validitätstestung durch die Überprüfung von verschiedenen Tests könnte die Validität eines Tests vielleicht eher bestätigen oder in Frage stellen.

Zu der vorliegenden Studie wäre zum Beispiel der Diaphragmatest mit Daumenpalpation unter dem Rippenbogen in Rückenlage eine interessante Ergänzung. Dieser Test berücksichtigt die Kuppelform des Diaphragmas und gibt ebenfalls Aufschluss über die Spannkraft des Diaphragmas. Die Testergebnisse der beiden Tests würden miteinander verglichen werden und falls eine Übereinstimmung herrscht, könnte man davon ausgehen, dass beide Tests valide sind.

Allgemein sei noch angeführt, dass ein zuverlässiger Test noch keine Relevanz hinsichtlich des osteopathischen Behandlungserfolgs hat. Ein nächster Studienschritt nach einer positiven Reliabilitätstestung wäre, eine Verbindung zwischen einem zuverlässigen Test und dessen Relevanz für die Behandlung und Verbesserung des Zustandes eines Patienten nachzuweisen.

Zu erwähnen ist auch der eigene Nutzen mit der vermehrten praktischen und literarischen Beschäftigung an einem Test, was zu einer vergrößerten Genauigkeit beim praktischen Testen an sich führt. Des Weiteren wird das Hinterfragen geschult, warum genau dieser Test in diesem Zusammenhang angewendet und wie dieser Test daraufhin interpretiert wird.

Intuition beruht nicht nur auf das "Bauchgefühl", sondern auch auf Erfahrung und Wissen, so kann sie durch Reflexion gestärkt werden (Sidler 2012). Eine Studie dieser Art kann die eigenen Behandlungstechniken in Frage stellen, Denkmuster in der Diagnosestellung neu beleuchten und so dazu beitragen, Intuition und Wissen in der osteopathischen Vorgangsweise optimal ergänzen. Retrospektive Fallbeurteilungen, Supervisionen oder das Schreiben einer Masterthese schulen das Wissen, decken eigene Denkmuster auf und verhelfen so zu einer Kombination von bewusster wie unbewusster Denkarbeit, also zu einer fundierten Intuition.

6. Schlussfolgerung

Die Untersucher kommen bei der Interrater Reliabilität signifikant seltener, als durch Zufall zu erwarten wäre, zu den gleichen Ergebnissen. Es gibt signifikante systematische Unterschiede zwischen den Testern. Untersucher 1 bevorzugt im Testergebnis die linke Körperseite, Untersucher 2 die rechte. Bei der Test-Retest Reliabilität hat nur Tester 1 eine mittelmäßig signifikante Übereinstimmung.

Aus dem Ergebnis kann herausgelesen werden, dass die Untersucher möglicherweise unterschiedlich auf die Spannungen jeder Körperseite der Probanden reagieren, also die Nuancen anders wahrnehmen und interpretieren.

Da die beiden Untersucher in den Kappawerten bei der Test-Retest Reliabilität unterschiedlich abschneiden, ist im Schlussfolgern das Abstrahieren der Fakten nur ein Mutmaßen.

Eine größere Anzahl an Untersuchern würde ein Verallgemeinern auf die Gesamtheit der Osteopathen und auch das Filtern von statistischen Ausreißern ermöglichen.

Der von den Untersuchern festgestellte erhöhte Grundtonus der Probanden könnte den palpatorischen Test erschwert haben. Zum Senken des Grundtonus der Probanden hätte vielleicht eine geregelte Testung alleinig in der Einatmung beitragen können. Weiters hätte eventuell das Aufheben der Blindierung die unbewusste Spannung um die Körpermitte der Probanden gesenkt.

In der normalen Untersuchungssituation in der Praxis verhilft die vorangegangene Anamnese und der an den Patienten angepasste Untersuchungsrythmus dem Patienten zu einer Entspannung, wo mit geringen Bewegungsamplituden Spannungszentren im Körper aufgespürt werden können. Eine Studie mit Aneinanderreihung von verschiedenen spezifischen Tests würde vielleicht mehr Rückschlüsse für die Diagnosestellung zulassen und praxisnaher sein, was allerdings keine klassische Reliabilitäts- sondern eine komplexe Validitätsstudie wäre.

Hätten die Untersucher beim Ergebnis anstatt zwischen "gleich, rechts und links" nur zwischen "rechts und links" gewählt, wäre vielleicht bei Grenzfällen die Entscheidung leichter gefallen und hätte das intuitive Gefühl mehr die Oberhand gehabt (was bei einem Test mit geringem Bewegungsimpuls von Vorteil sein kann).

Obwohl durch genaues Recherchieren, eine Probeuntersuchung und eine Einschulung der Untersucher Fehlerquellen gering zu halten versucht wurden, können letzten Endes folgende Vorschläge zur optimierten Durchführung angeführt werden:

Nach der Einschulung der Untersucher wäre eine Gewöhnungsphase der Testausführung in der Größenordnung eines Monats angebracht. Die Einschulung sollte die Griffhaltung individuell belassen, jedoch fundiertes Hintergrundwissen über die Diaphragmaebene vor

allem für die Intention der Palpationstiefe schaffen. Sinnvoll wäre dann eine Überprüfung der Testausführung mit einer anschließenden Angleichung der Palpationstiefe am Versuchsleiter und erneut einer einmonatigen Automatisierungsphase. Eine Probedurchführung mit statistischer Auswertung würde zusätzliche Fehlerquellen im Vorfeld aufdecken können.

Bei der Auswahl der Probanden würden meiner Meinung nach gesunde Probanden ausreichen. 30 Probanden und sechs Untersucher sind meines Erachtens für ein gut verwertbares Ergebnis geeignet.

Als interessante Ergänzung, und vielleicht auch zur Einstimmung für die Entspannung der Probanden, wäre der Diaphragmatest mit Daumenpalpation unter dem Rippenbogen als Referenzwert zur eigentlichen Untersuchung geeignet.

Der Spannungstest des abdominalen Diaphragmas ist für mich und auch für beide Untersucher in dieser Studie trotzdem ein guter Überblickstest in der Praxis, um die hypertone Seite im Patienten herauszukristallisieren. Der darauffolgende Behandlungserfolg spricht für die Kombination mit dem Diaphragmatest mit Daumenpalpation und eines dynamischen Diaphragmatests mit translatorisch und rotatorisch induzierter Bewegung. Die Reliabilitätsprüfung ausgeführt wie in dieser Masterthese ist für die Osteopathie in Frage zu stellen. Die individuelle Freiheit des Osteopathen bei der Testausführung und -interpretation ist im Zusammenhang der Komplexheit des menschlichen Körpers und Geistes vom Patient und Untersucher erfolgsversprechend.

7. Abkürzungsverzeichnis

BMI = Bodymassindex

C = Halswirbel

Cx = Halswirbelsäulenbereich

EMG = Elektromyogramm

G = gleich/beidseits

ID = Identifikationsnummer der Reihenfolge der Probanden bei der Untersuchung

L = links

Lx = Lendenwirbelsäulenbereich

m = männlich

M. = Musculus

M = Mittelwert

N. = Nervus

Nr. = zwecks Anonymisierung verwendete Kennzeichennummer für einen Probanden

Pat. = Patient

R = rechts

SD = Standardabweichung

Sicher* = von zumindest einem Tester als "sicher" eingeschätztes Ergebnis

Sicher** = vom jeweiligen Tester als "sicher" eingeschätztes Ergebnis

TFL = Thorakolumbale Faszie

T1 = Tester 1

T2 = Tester 2

TLÜ = Bereich des thorakolumbalen Übergangs

Tx = Brustwirbelsäulenbereich

U = Untersuchung

Unters. = Untersuchung

w = weiblich

+ = Übereinstimmung

- = keine Übereinstimmung

Fett geschriebene Angabe = sicheres Ergebnis

8. Abbildungsverzeichnis

Abb.1: Schematische Darstellung eines kollagenen Netzwerks im entspannten Normalzustand (a), entspannt mit pathologischen Crosslinks (b), im belasteten Normalzustand (c), und belastet mit pathologischen Crosslinks (d) (Van den Berg, 1999, S.179)

Abb.2: Zirkuläres Modell für evidence practised medicine (Walach, 2006)

Abb.3: Abdominales Diaphragma (Schünke, 2005, S.147)

Abb.4: Diaphragma im transversalen Überblick (Schünke, 2005, S.134)

Abb.5: 3D-Rekonstruktion des Diaphragmas, Ansicht von ventral in maximaler Einatmung bzw. Ausatmung (Meert, 2007, S.136)

Abb.6: Kreuzung der Körperseite durch M. Obliquus Externus und Adduktoren, M. Obliquus Internus und M. Gluteus Medius (Myers, 2004, S.180).

Abb.7: Fixierter Zwerchfellhochstand bzw. Fixierter Zwerchfelltiefstand (Meert, 2007, S.141)

Abb.8: Kompensationsfähige und nicht kompensationsfähige Muster (Pope, 2003)

Abb.9: Test der abdominalen Diaphragmaspannung (Meert, 2007, S.324)

Abb.10: Blindierung für Proband und Untersucher

Abb.11: Schematischer Überblick zum Untersuchungsablauf

Abb.12: Ausführung des Diaphragmentests bei der Studie

9. Tabellenverzeichnis

Tab.1: Ergebnisse der Probedurchführung 05.09.2012

Tab.2: Textliche Interpretation des Kappa-Index

Tab.3: Beobachtbare und erwartete Übereinstimmung bei der Interrater Reliabilität

Tab.4: Beobachtbare und erwartete Übereinstimmung bei der Interrater Reliabilität der sicheren Ergebnisse

Tab.5: Beobachtbare und erwartete Häufigkeit der Test-Rest Reliabilität beim Tester 1

Tab.6: Beobachtbare und erwartete Häufigkeit der Test-Rest Reliabilität beim Tester 2

Tab.7: Übereinstimmung der Schmerzangabe des Patienten mit dem Ergebnis von Tester 1

Tab.8: Übereinstimmung der Schmerzangabe des Patienten mit dem Ergebnis von Tester 2

Tab.9: Übereinstimmung der Rechts oder Links Schmerzangabe des Patienten mit dem Ergebnis von Tester 1

Tab.10: Übereinstimmung der Rechts oder Links Schmerzangabe des Patienten mit dem Ergebnis von Tester 1

Tab.11: Beobachtete und erwartete Übereinstimmung der Untersuchungsreihenfolge Tester 2 -Tester 1 (Interrater Reliabilität)

Tab.12: Beobachtete und erwartete Übereinstimmung der Untersuchungsreihenfolge Tester 1 - Tester 1 (Interrater Reliabilität)

Tab.13: Gegenüberstellung der Häufigkeit der Untersuchungsergebnisse in den beiden Räumen von den beiden Testern

Tab.14: Gegenüberstellung der Häufigkeit der Untersuchungsergebnisse der beiden Tester im Testdurchgang 1

Tab.15: Gegenüberstellung der Häufigkeit der von den beiden Testern als „sicheres“ Ergebnis bezeichneten Untersuchungsergebnisse

10. Literaturverzeichnis

Angele C, Derndinger J. (2011): Diagnose-Medizinlexikon. Verfügbar unter: <http://www.imedo.de/medizinlexikon/diagnose> [03.12.2011].

Bengaard K, Bogue RJ, Crow WT. (2012): Reliability of diagnosis of somatic dysfunction among osteopathic physicians and medical students. *Osteopathic Family Physician* 4(1):2-7.

Benjamin M. (2009): The fascia of the limbs and back-a review. *Journal of Anatomy* 214:1-18.

Brazzo M. (2004): Viszerale Automobilisation-Osteopathie für die inneren Organe. München. Elsevier.

Chen CS, Ingber DE. (2008): Tensegrity und Mechanoregulation: Vom Skelett zum Zytoskelett. *Osteopathische Medizin* 9(4):4-16.

Clare HA, Adams R, Maher CG. (2005): Reliability of McKenzie classification of patients with cervical or lumbar pain. *Journal of Manipulative and Physiological Therapeutics* 28(2):122-127.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20: 37-46.

Croibier A. (2006): Diagnostik in der Osteopathie. München. Elsevier.

Danto JB. (2003): review of integrated neuromusculoskeletal release and the novel application of a segmental anterior/posterior approach in the thoracic, lumbar and sacral regions. *Journal of the American Osteopathic Association* 103(12):583-596.

Degenhardt BF, Johnson JC, Snider KT, Snider EJ. (2010): Maintenance and Improvement of Interobserver Reliability of Osteopathic Palpatory Tests Over a 4-Month Period. *Journal of the American Osteopathic Association* 110(10):579-586.

Defeo G, Hicks L. (1993): A description of the common compensatory pattern in relationship to osteopathic postural examination. *American Academy Osteopathic Journal*.

Feipel V, Rondelet B, Le Pallec JP, De Witte O, Rooze M. (1999): The use of disharmonic motion curves in problems of the cervical Spine. *International Orthopaedics* 23(4): 205-209.

Finet G, Williame C. (2007): Deltadyn - Articular - Visceral. <http://www.deltadyn.be/visceral/ger/einleitung.htm> [28.03.2013].

Fjellner A, Bexander C, Faleji R, Strender LE. (1999): Interexaminer reliability in physical examination of the cervical spine. *Journal of Manipulative and Physiological Therapeutics* 22: 511-516.

Fritz, JM, Wainner RS. (2001): Examining Diagnostic Tests: An Evidence-based Perspective. *Physical Therapy* 81(9):1546-1564.

Glover JC, Clem A, DeStefano L, Devine W, Ehrenfreuchter W, Essig-Beatty D, Ferrell H, Fotopolis T, Gilliar W, Heinking K, Hendryx J, Hrubby R, Jones J, Kappler R, Kerger S, Litman R, Lockwood M, Mason D, Nicholas E, Pim K, Rennie P, Sandhouse M, Shaw H, Thompson G, Treffer K. (2006): *Glossary of Osteopathic Terminology Usage Guide*. California. The Educational Council on Osteopathic Principles and the American Association of Colleges of Osteopathic Medicine.

Greenman PE. (2005): *Lehrbuch der osteopathischen Medizin*. Stuttgart. Haug.

Guimberteau JC, Rigall JS, Panconi B, Bouleau R, Moutan P, Bakhach J. (2008): Die Gleitfähigkeit subkutaner Strukturen beim Menschen – eine Einführung. *Osteopathische Medizin* 9(1):4-16.

Halma KD, Degenhardt BF, Snider KT, Johnson JC, Flaim MS, Bradshaw D. (2008): Intraobserver Reliability of Cranial Strain Patterns as Evaluated by Osteopathic Physicians: A Pilot Study. *Journal of the American Osteopathic Association* 108(9):493-502.

Haneline MT, 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.

Hartmann C, Hrsg. (2005): *Das große Still Kompendium*. Unterwössen. Jolandos.

Helsmoortel J. (2002): *Lehrbuch der visceralen Osteopathie*. Stuttgart. Thieme.

Hodges PW, Gandevia SC. (2000): Activation of the human diaphragm during a repetitive postural task. *Journal of Physiology* 522(1):165-176.

Johnston WL. (1982): Interexaminer reliability studies: spanning a gap in medical research-Luisa Burns Memorial lecture. *Journal of the American Osteopathic Association* 81(12):819-829.

- Kary DJ. (2009): The Transversus Thoracis Muscle in humans- diagnosis and treatment of associated pathology: an osteopathic perspective. *Journal of the American Osteopathic Association* 19(4):21-30.
- Kilpikoski S, Airaksinen O, Kankaanpää M, Leminen P, Videman T, Alen M. (2002): Interexaminer reliability of low back pain assessment using the McKenzie method. *Spine* 27(8):207-214.
- Klein P, Sommerfeld P. (2004): *Biomechanik der menschlichen Gelenke–Grundlagen, Becken untere Extremität*. München. Elsevier.
- Krause R. (2008): *Palpation – Wahrnehmung – Heilung*. Stuttgart. Sonntagverlag.
- Kuchera ML. (2007): Applying osteopathic principles to formulate treatment for patient with chronic pain. *Journal of the American Osteopathic Association* 107(11)6:28-38
- Landis JR, Koch GG. (1977): The measurement of observer agreement for categorical data. *Biometrics* 33: 159-174.
- Lancaster DG, Crow WT. (2006): Osteopathic manipulative treatment of a 26-year-old woman with Bell's palsy. *Journal of the American Osteopathic Association* 106(5):285-290.
- Lavelle JM. (2012): Osteopathic manipulative treatment in pregnant women. *Journal of the American Osteopathic Association* 112(6):343-346.
- Leonhardt H, Tillmann B, Töndury G, Zilles K. (1987): *Rauber/Kopsch- Anatomie des Menschen. Bewegungsapparat, Band I*. Stuttgart. Thieme.
- Liem T. (2005): *Kraniosacrale Osteopathie. Ein praktisches Lehrbuch*. Stuttgart. Hippokrates.
- Liem T. (2006): *Morphodynamik in der Osteopathie*. Stuttgart. Hippokrates.
- Liem T, Hilbrecht H, Schmidt T. (2012): Osteopathie und Wissenschaft. *Osteopathische Medizin* 13(1):11-17.
- Mayer-Fally E. (2010): *Klinische Osteopathie - Bestandsaufnahme der klinischen Arbeit in der Osteopathie*. Verfügbar unter: <http://studenten.wso.at/index.php/212-neue-skripten-im-e-learning-bereich> [26.03.2013].
- Meert GF. (2007): *Das venöse und lymphatische System aus osteopathischer Sicht*. München. Elsevier.

- Moran RW, Gibbons P. (2001): Intraexaminer and interexaminer reliability for palpation of the cranial rhythmic impulse at the head and sacrum. *Journal of manipulative and physiological therapeutics* 24(3):183-190.
- Münch G. (2011): *Manuelle Stimmtherapie (MST), eine Therapie, die berührt*. Idstein. Schulz-Kirchner Verlag.
- Myers TW. (2004): *Anatomy Trains - Myofasziale Meridiane*. München. Elsevier.
- Viereck GS. (1929, 26. Oktober): What life means to Einstein - An interview by George Sylvester Viereck. *The saturday evening post*. Verfügbar unter: http://www.saturdayeveningpost.com/wp-content/uploads/satevepost/what_life_means_to_einstein.pdf [Stand: 26.03.2013].
- Norton JM. (1996): A challenge to the concept of craniosacral interaction. Verfügbar unter: <http://faculty.une.edu/com/jnorton/challenge.htm> [Stand: 04.03.2013].
- Olivo SA, Macedo LG, Gadotti IA, Fuentes J, Stanton T, Magee DJ. (2008): Scales to assess the quality of randomized controlled trials: a systematic review. *Physical therapy* 88(2):156-175.
- Oschman JL. (2002): Clinical aspects of biological fields: an introduction for health care professionals. *Journal of bodywork and movement therapies* 6(2):117-125.
- Paoletti S. (2001): *Faszien. Anatomie, Strukturen, Techniken, Spezielle Osteopathie*. München. Elsevier.
- Parodi RZ, Chauvigny de Blot P, Rickards LD, Renard EO. (2009): Cranial palpation pressures used by Osteopathy students: Effects of standadized protocol training. *Journal of American Osteopathic Association* 109(2):79-85.
- Patijn J. (2004): *Reproducibility and validity studies of Diagnostic Procedures in Manual/Musculoskeletal Medicine. Protocol formats*. International Federation for Manual/Musculoskeletal Medicine. Maastricht.
- Pflüger C. (2010): Tensegrity in der Osteopathie. *Osteopathische Medizin* 11(3):10-13.
- Pickering M, Jones JFX. (2002): The diaphragm: two physiological muscles in one. *Journal of Anatomy* 201:305-312.

Pope RE. (2003): The common compensatory pattern: its origin and relationship to the postural model. American Academy Osteopathic journal. Verfügbar unter: google scholar: RE Pope - Am Acad Osteopath J, 2003 - virginiasmiles.com [Stand: 20.02.2013].

R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rittler M. (2010): Ist das Global Listening reliabel? Studie zur Intrarater- und Interrater-Reliabilität des Global Listenings. Masterthese. Universität Krems.

Schmidt K, Liem T. (2012): Interview mit Jaap van der Wal über die Emryologie und ihre Bedeutung für die Osteopathie. Osteopathische Medizin 12(2):13-17.

Schünke M, Schulte E, Schumacher U, Voll M, Wesker K. (2005): Prometheus- Lernatlas der Anatomie. Hals und Innere Organe, Band II. Stuttgart. Thieme.

Sergueef N, Greer MA, Nelson KE, Glonek T. (2011): Kranialer rhythmischer Impuls: Ist die palpierete Frequenz abhängig von der Berufserfahrung der Untersucher? Osteopathische Medizin 12(4):4-11.

Sidler S (2012): Wie denken die Finger? Unbewusstes Denken als eine Grundlage der Intuition. Osteopathische Medizin 13(3):4-9.

Snider KT, Johnson JC, Snider EJ, Degenhardt BF. (2008): Increased incidence and severity of somatic dysfunction in subjects with chronic low back pain. Journal of the American Osteopathic Association 108(8):372-378.

Sommerfeld P, Kaider A, Klein P. (2004): Inter- and intraexaminer reliability in palpation of the "primary respiratory mechanism" within the "cranial concept". Manual therapy 9(1):22-29.

Stovall BA, Kumar S. (2010): Reliability of Bony Anatomic Landmark Asymmetry Assessment in the Lumbopelvic Region: Application to Osteopathic Medical Education. Review. Journal of the American Osteopathic Association 110(11):667-674.

Ünal Ö, Arslan H, Uzun K, Özbay B, Sakaraya ME. (2000): Evaluation of diaphragmatic movement with MR fluoroscopy in chronic obstructive pulmonary disease. Clinical Imaging 24(6):347-350.

Van den Berg F, Cabri J. (1999): Angewandte Physiologie 1 - Das Bindegewebe des Bewegungsapparates verstehen und beeinflussen. Stuttgart, Thieme.

Van den Berg F, Cabri J, Elvey B, Gosselink R, Haas HJ, Heesen G, Horst R, Van Kampen M, Oettmeier R, Reybrouck T, Schöttker-Königer T, Sinz H, Slater H, Steverding M, Thacker MA, Watson T, Wilke M. (2001): *Angewandte Physiologie 3 - Therapie, Training, Tests*. Stuttgart. Thieme.

Walach H, Falkenberg T, Fonnebo V, Lewith G, Jonas WB. (2006): Circular instead hierarchical: methodological principles for the evaluation of complex interventions. Verfügbar unter: <http://www.biomedcentral.com/1471-2288/6/29> [Stand: 20.02.2013].

Whitelaw WA. (1987): Shape and size of the human diaphragm in vivo. *Journal of Applied Physiology* 62(1):180-186.

Wisser R. (1995): *Karl Jaspers: Philosophie in der Bewährung, Vorträge und Aufsätze*. Würzburg. Königshausen&Neumann GmbH.

Wolke N. (2009): *The three leaves of Glenard-Examination of a functional aspect- An Interexaminer- and Intraexaminer Reliability Study*. Masterthese. Universität Krems.

11. Anhang

Ärzteinformation

Herzlichen Dank, dass Sie Patienten zu meiner Studie vermitteln!

Zeit: Freitag, 19. 10. 2012, 18.00h - 19.30h

Ort: Praxis Physiotherapie Traunsee, 4810 Gmunden
Platz der Sudetendeutschen 8

Studienleiterin: Sonja Gruber, Osteopathin und Physiotherapeutin

Tel.: 0699/81329542

Einschlusskriterien: 20 Probanden zwischen 18 und 75 Lebensjahren mit einem subjektiv stabilen gesundheitlichen Zustand und 20 Probanden mit Schmerzen im thorakalen, thorakolumbalen oder lumbalen Bereich mindestens fünf Tage in der Woche seit mindestens drei Monaten.

Ausschlusskriterien: Personen mit akutem Geschehen im Thorax und Abdomen (Pathologien, Frakturen) und schwangere Frauen.

Für die Erstellung meiner Masterthese führe ich eine Studie durch.

Dabei überprüfe ich einen Test der osteopathischen Untersuchung auf seine Zuverlässigkeit.

Es handelt sich um die Testung der abdominalen Diaphragmaspannung im Seitenvergleich.

Die Untersucher sind Osteopathen mit langjähriger Berufserfahrung.

Der Proband liegt in Rückenlage unterstützt mit Knierolle und Nackenpolster. Der Osteopath legt seine Hände beidseits auf den Rippenbogen des Probanden von lateral kommend. Der Brustkorb des Probanden wird passiv und leicht, in geringer Bewegungsamplitude, von der einen zur anderen Seite transversal verschoben. Für die Untersuchung ist das Freimachen des Bauchbereichs vom Rippenbogen bis zum Beckenknochen erforderlich. Die Probanden sind mit Leintüchern vor dem Untersucher abgedeckt. Eine Testung dauert einige Sekunden. Diese Testung wird viermal durchgeführt.

Für das leibliche Wohl ist in Form eines Buffets gesorgt.

Ich bitte Sie, die Patienten zu erinnern, mich bei Teilnahme der Studie alsbald telefonisch zu kontaktieren, damit ich die Probandenzahl abschätzen kann und auch wirklich 20 Patienten zusammen bekomme!

Vielen Dank und freundliche Grüße

Sonja Gruber

Probandeninformation

Herzlichen Dank, dass Sie bei meiner Studie teilnehmen!

Zeit: Freitag, 19. 10. 2012, 16.00-18.00h

Ort: Praxis Physiotherapie Traunsee, 4810 Gmunden
Platz der Sudetendeutschen 8

Studienleiterin: Sonja Gruber, Osteopathin und Physiotherapeutin

Tel.: 0699/81329542

Für die Erstellung meiner Masterthese führe ich eine Studie durch.

Dabei überprüfe ich einen Test der osteopathischen Untersuchung auf seine Zuverlässigkeit.

Die Untersucher sind Osteopathen mit langjähriger Berufserfahrung.

In Rückenlage mit Knierolle und Nackenpolster als Unterstützung werden Sie auf Höhe des Rippenbogens sanft in einem geringen Bewegungsausmaß nach rechts und links geschaukelt. Das Freimachen des Bauchbereichs vom Rippenbogen bis zum Becken ist erforderlich. Sie sind mit Leintüchern vor dem Untersucher abgedeckt.

Eine Testung dauert einige Sekunden und wird viermal durchgeführt.

Ich bitte Sie, an diesem Tag kein Parfum zu verwenden und kurzärmelig und mit bequemer Kleidung zu erscheinen.

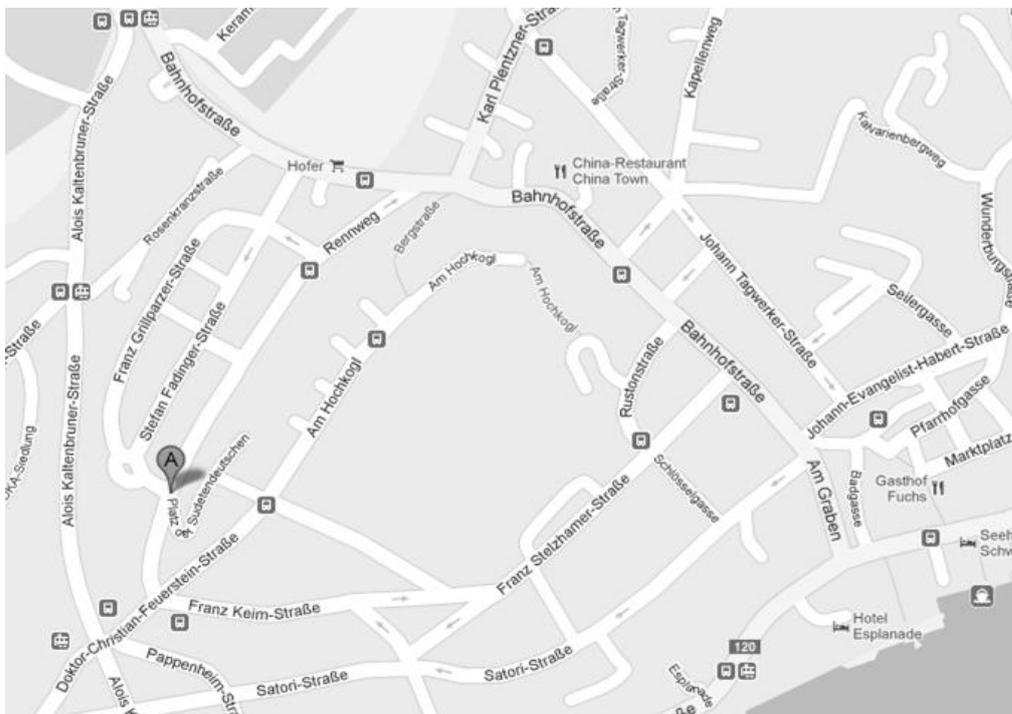
Da 40 Personen an dieser Studie teilnehmen, sind wegen der organisatorischen Abwicklung zwei Stunden als Studiendauer geplant.

Für das leibliche Wohl ist in Form eines Buffets gesorgt.

Ich bitte Sie um baldige telefonische Anmeldung und um ein verlässliches Erscheinen um 16.00h, da ich bei geringerer Personenanzahl meine Daten nicht verwerten kann.

Freundliche Grüße und bis bald!

Sonja Gruber



Parkmöglichkeiten:

- ◆ entlang der Stefan Fadinger Straße
- ◆ entlang vom Rennweg (Einbahnstraßen!)
- ◆ und am Platz der Sudetendeutschen

Fragebogen für die Probanden**Informationsblatt zur Studie am 19.10.2012 für die Masterthese im Lehrgang für Osteopathie an der Universität Krems**

Ich freue mich sehr über Ihre Teilnahme. Herzlichen Dank!

Damit ich die Daten gut auswerten kann, brauch ich noch einige Informationen über Sie:

Name:

Losungsnummer:

Alter:

Gewicht:

Haben Sie Schmerzen im Rückenbereich, mindestens 5 Tage in der Woche seit mindestens 3 Monaten?

Ja

Nein

Wenn ja: Wo sind die Schmerzen?

RECHTS

LINKS

BEIDSEITS

Brustwirbelsäulenbereich

Übergangsbereich

Lendenwirbelsäulenbereich

Nun noch einige Informationen zum Ablauf:

Sie bekommen eine Nummer, die sichtbar im Schulterbereich angebracht werden sollte. Mit dieser Nummer wird die Reihenfolge der getesteten Personen bestimmt und ob sie zweimal oder viermal untersucht werden. Sie werden zum gegebenen Zeitpunkt zum Untersuchungsort geleitet.

Der Test findet in Rückenlage statt. Dazu ist das Freimachen des Bauchbereichs vom Rippenbogen bis zum Becken erforderlich. In Rückenlage mit Knierolle und Nackenpolster als Unterstützung werden Sie auf Höhe des Rippenbogens sanft in einem geringen Bewegungsausmaß nach rechts und links geschaukelt. Eine Testung dauert nur einige Sekunden. Für den Test ist es wichtig, dass Sie der Untersucher nicht sieht. Darum sind Sie mit Leintüchern abgedeckt. Die Untersucher sind Osteopathen mit langjähriger Berufserfahrung.

Während dem Test ist ein entspanntes Liegen mit gleichmäßigem Atmen erwünscht.

Wenn Ihre Untersuchungen beendet sind, darf ich zum Buffet einladen.

Feedbackbogen Probanden

Herzlichen Dank für Ihre Teilnahme!

Abschließend habe ich noch ein paar Fragen an Sie:

1. Gab es beim Test Unterschiede bei der Griffhaltung des Untersuchers?

Ja Nein

Anmerkung:

2. Gab es beim Test Unterschiede in der Druckstärke beim Griff des Untersuchers?

Ja Nein

Anmerkung:

3. Hatten sie das Gefühl vom gleichen Osteopathen untersucht worden zu sein?

Ja Nein

Anmerkung:

4. Wie war Ihr Befinden während dem Test?

Wohlbefinden Schmerzempfinden

Anmerkung:

5. Wie ist Ihr Befinden nach dem Test?

Wohlbefinden Schmerzempfinden

Anmerkung:

6. Sonstige Anmerkungen:

Probandentabelle

Nr.	Alter	m/w	Gesund	Schmerz	r/g/l	Tx/TLÜ/Lx	Gewicht	Größe
1	37	W	X				70	176
2	30	W	X				66	170
3	75	M	X				76	167
4	48	M	X				80	184
5	39	M	X				77	192
6	41	W	X				66	165
7	37	W	X				60	170
8	59	W	X				57	169
9	67	M	X				95	185
10	46	W	X				68	174
11	69	W	X				52	156
12	64	W	X				66	170
13	52	M	X				80	180
14	52	M	X				82	177
15	64	W	X				61	164
16	45	W	X				65	178
17	40	W	X				58	167
18	70	W	X				67	167
19	44	M	X				86	186
20	33	W	X				60	170
42	69	M	X				66	174
43	40	M	X				96	188
21	65	W		X	r	Lx	60	165
22	70	M		X	l	Lx	84	182
23	71	M		X	l	Lx	92	176
24	43	W		X	l	Lx	55	166
25	39	M		X	g	Lx	86	191
26	60	W		X	l	Tx	69	169
27	59	W		X	l	Tx	52	162
28	67	M		X	r	Lx	76	173
29	61	W		X	g	TLÜ,Lx	78	161
30	74	M		X	g	Lx	86	178
31	42	W		X	g	Tx,Lx	70	166
32	72	W		X	l	Tx	71	174
33	75	M		X	g	Lx	73	168
34	63	M		X	r	TlÜ	80	174
35	66	W		X	l	Lx	70	160
36	69	M		X	l	Lx	95	176
37	66	W		X	g	Lx	83	165
38	59	W		X	r	Lx	59	160
39	67	M		X	r	Lx	82	177
40	66	M		X	g	Lx	83	183
41	34	W		X	g,mehr l	Tx	70	161

Originale Eintragungen der Untersuchungsergebnisse



RAUM A **GEORG (Tester 1)**

Diaphragmaspannung (falls Ergebnis sicher: EINKREISEN)

Nr.	Proband	RECHTS	GLEICH	LINKS
1 A	33			⊗
2 A	24		×	
3 A	42			×
4 A	4	×		
5 A	2			×
6 A	4	×		
7 A	22			⊗
8 A	30			⊗
9 A	31			×
10 A	19		×	
11 A	23	×		
12 A	1			×
13 A	8		×	
14 A	23	×		
15 A	26			⊗
16 A	8			×
17 A	20			×
18 A	2	×		
19 A	25	×		
20 A	6	⊗		
21 A	36			⊗
22 A	32	×		
23 A	28	×		
24 A	28	×		
25 A	5			×
26 A	6	×		
27 A	38			×
28 A	25	×		
29 A	17	×		
30 A	34			×
31 A	35	⊗		

Jlx

WOLFGANG 1.


RAUM B**GEORG (Tester 1)****Diaphragmaspannung (falls Ergebnis sicher: EINKREISEN)**

Nr.	Proband	RECHTS	GLEICH	LINKS
1 B	16	X		
2 B	3			X
3 B	43		X	
4 B	18	X		
5 B	5			⊗
6 B	24			X
7 B	27			⊗
8 B	3			X
9 B	40			X
10 B	1			⊗
11 B	10			X
12 B	15			X
13 B	14			X
14 B	22			X
15 B	11	X		
16 B	21			⊗
17 B	27			X
18 B	21			X
19 B	10			X
20 B	29		X	
21 B	29		X	
22 B	30			⊗
23 B	7			X
24 B	26			X
25 B	37	X		
26 B	39	X		
27 B	13		X	
28 B	12			X
29 B	9	X		
30 B	9		X	
31 B	41		X	
32 B	7			X

Elke

↑ GEORG (1.)

RAUM A

WOLFGANG (Tester 2)

Diaphragmaspannung (falls Ergebnis sicher: EINKREISEN)

Nr.	Proband	RECHTS	GLEICH	LINKS
1 A	33			⊗
2 A	24			X
3 A	42	X		
4 A	4			X
5 A	2	⊗		
6 A	4	X		
7 A	22	X		
8 A	30	X		
9 A	31			X
10 A	19			X
11 A	23		X	
12 A	1	X		
13 A	8			X
14 A	23			X
15 A	26		X	
16 A	8	X		
17 A	20	X		
18 A	2	⊗		
19 A	25		X	
20 A	6			X
21 A	36		X	
22 A	32	X		
23 A	28	X		
24 A	28	X		
25 A	5	X		
26 A	6			X
27 A	38	⊗		
28 A	25	X		
29 A	17		X	
30 A	34	⊗		
31 A	35			X

Jesc

WOLFGANG ①

RAUM B

WOLFGANG (Tester 2)

Diaphragmaspannung (falls Ergebnis sicher: EINKREISEN)

Nr.	Proband	RECHTS	GLEICH	LINKS
1 B	16			X
2 B	3	X		
3 B	43			X
4 B	18		X	
5 B	5	X		
6 B	24	X		
7 B	27		X	
8 B	3	X		
9 B	40		X	
10 B	1	X		
11 B	10	X		
12 B	15	X		
13 B	14	X		
14 B	22			X
15 B	11			X
16 B	21	X		
17 B	27	X		
18 B	21	X		
19 B	10		X	
20 B	29	⊗		
21 B	29	X		
22 B	30	⊗		
23 B	7	X		
24 B	26			X
25 B	37			X
26 B	39			X
27 B	13	⊗		
28 B	12		X	
29 B	9	X		
30 B	9		X	
31 B	41	X		
32 B	7			X

Reihenfolge gesamt

Gesamt: Reihenfolge				
Reihenfolge	Tester 1		Tester 2	
	ID	Ergebnis	ID	Ergebnis
1	33	l	16	l
2	16	r	33	l
3	24	g	3	r
4	3	l	24	l
5	42	l	43	l
6	43	g	42	r
7	4	r	18	g
8	18	r	4	l
9	2	l	5	r
10	5	l	2	r
11	4	r	24	r
12	24	l	4	r
13	22	l	27	g
14	27	l	22	r
15	30	l	3	r
16	3	l	30	r
17	31	l	40	g
18	40	l	31	l
19	19	g	1	r
20	1	l	19	l
21	23	r	10	r
22	10	l	23	g
23	1	l	15	r
24	15	l	1	r
25	8	g	14	r
26	14	l	8	l
27	23	r	22	l
28	22	l	23	l
29	26	l	11	l
30	11	r	26	g
31	8	l	21	r
32	21	l	8	r
33	20	l	27	r
34	27	l	20	r
35	2	r	21	r
36	21	l	2	r
37	25	r	10	g
38	10	l	25	g
39	6	r	29	r
40	29	g	6	l
41	36	l	29	r
42	29	g	36	g
43	32	r	30	r
44	30	l	32	r
45	28	r	7	r
46	7	l	28	r
47	28	r	26	l
48	26	l	28	r
49	5	l	37	l
50	37	r	5	r
51	6	r	39	l
52	39	r	6	l
53	38	l	13	r
54	13	g	38	r
55	25	r	12	g
56	12	l	25	r
57	17	r	9	r
58	9	r	17	g
59	34	l	9	g
60	9	g	34	r
61	35	r	41	r
62	41	g	35	l
63	7	l	7	l

Reihenfolge Proband

Reihenfolge der Untersuchungen am Probanden								
ID	1. Untersuchung		2. Untersuchung		3. Untersuchung		4. Untersuchung	
	Tester	Ergebnis	Tester	Ergebnis	Tester	Ergebnis	Tester	Ergebnis
1	2	r	1	l	1	l	2	r
2	1	l	2	r	1	r	2	r
3	2	r	1	l	2	r	1	l
4	1	r	2	l	1	r	2	r
5	2	r	1	l	1	r	2	r
6	1	r	2	l	1	r	2	l
7	2	r	1	l	2	l	1	l
8	1	g	2	l	1	l	2	r
9	2	r	1	r	2	l	1	g
10	2	r	1	l	2	g	1	l
11	2	l	1	r				
12	2	g	1	l				
13	2	r	1	g				
14	2	r	1	l				
15	2	r	1	l				
16	2	l	1	r				
17	1	r	2	g				
18	2	g	1	r				
19	1	g	2	l				
20	1	l	2	r				
21	2	r	1	l	2	r	1	l
22	1	l	2	r	2	l	1	l
23	1	r	2	g	1	r	2	l
24	1	g	2	l	2	r	1	l
25	1	r	2	g	1	r	2	r
26	1	l	2	g	2	l	1	l
27	2	g	1	l	2	r	1	l
28	1	r	2	r	1	r	2	r
29	2	r	1	g	2	r	1	g
30	1	l	2	r	2	r	1	l
31	1	l	2	l				
32	1	r	2	r				
33	1	l	2	l				
34	1	l	2	r				
35	1	r	2	l				
36	1	l	2	g				
37	2	l	1	r				
38	1	l	2	r				
39	2	l	1	r				
40	2	g	1	l				
41	2	r	1	g				
42	1	l	2	r				
43	2	l	1	g				

Interreliabilität im Testdurchgang 1

Interrater	Tester 1	Tester 2	Patient	
	Ergebnis	Ergebnis	Schmerz	erwartet
1	l	r		
2	l	r		
3	l	r		
4	r	l		
5	l	r		
6	r	l		
7	l	r		
8	g	l		
9	r	r		
10	l	r		
11	r	l		
12	l	g		
13	g	r		
14	l	r		
15	l	r		
16	r	l		
17	r	g		
18	r	g		
19	g	l		
20	l	r		
21	l	r	r	l
22	l	r	l	r
23	r	g	l	r
24	g	l	l	r
25	r	g	g	g
26	l	g	l	r
27	l	g	l	r
28	r	r	r	l
29	g	r	g	g
30	l	r	g	g
31	l	l	g	g
32	r	r	l	r
33	l	l	g	g
34	l	r	r	l
35	r	l	l	r
36	l	g	l	r
37	r	l	g	g
38	l	r	r	l
39	r	l	r	l
40	l	g	g	g
41	g	r	g	g
42	l	r		
43	g	l		

absolute Häufigkeiten				
l	22	13	8	5
r	14	21	5	8
g	7	9	8	8

Test- Retest Reliabilität im Testdurchgang 2

Test-Retest Reliabilität								
	Tester 1				Tester 2			
	Untersuchung 3		Untersuchung 4		Untersuchung 3		Untersuchung 4	
ID	Reihenfolge	Ergebnis	Reihenfolge	Ergebnis	Reihenfolge	Ergebnis	Reihenfolge	Ergebnis
1	10b	l	12a	l	10b	r	12a	r
2	5a	l	18a	r	5a	r	18a	r
3	2b	l	8b	l	2b	r	8b	r
4	4a	r	6a	r	4a	l	6a	r
5	5b	l	25a	r	5b	r	25a	r
6	20a	r	26a	r	20a	l	26a	l
7	23b	l	32b	l	23b	r	32b	l
8	13a	g	16a	l	13a	l	16a	r
9	29b	r	30b	g	29b	r	30b	l
10	11b	l	19b	l	11b	r	19b	g
21	16b	l	18b	l	16b	r	18b	r
22	7a	l	14b	l	7a	r	14b	l
23	11a	r	14a	r	11a	g	14a	l
24	2a	g	6b	l	2a	l	6b	r
25	19a	r	28a	r	19a	g	28a	r
26	15a	l	24b	l	15a	g	24b	l
27	7b	l	17b	l	7b	g	17b	r
28	23a	r	24a	r	23a	r	24a	r
29	20b	g	21b	g	20b	r	21b	r
30	8a	l	22b	l	8a	r	22b	r

absolute Häufigkeiten							
l		11		11		4	6
r		6		7		12	13
g		3		2		4	1

Untersuchungsreihenfolge von Tester 1 und 2

Untersuchung 1: Tester 2, Untersuchung 2: Tester 1

Interrater	Tester 1	Tester 2	Übereinstimmung
ID	Ergebnis U2	Ergebnis U1	T1/T2
1	l	r	-
3	l	r	-
5	l	r	-
7	l	r	-
9	r	r	+
10	l	r	-
11	r	l	-
12	l	g	-
13	g	r	-
14	l	r	-
15	l	r	-
16	r	l	-
18	r	g	-
21	l	r	-
27	l	g	-
29	g	r	-
37	r	l	-
39	r	l	-
40	l	g	-
41	g	r	-
43	g	l	-

absolute Häufigkeiten		
l	11	5
r	6	12
g	4	4
Summe:	21	21

absolute Häufigkeiten	
+	1
-	20
5%	

Untersuchung 1: Tester 1, Untersuchung 2: Tester 2

Interrater	Tester 1	Tester 2	Übereinstimmung
ID	Ergebnis U1	Ergebnis U2	T1/T2
2	l	r	-
4	r	l	-
6	r	l	-
8	g	l	-
17	r	g	-
19	g	l	-
20	l	r	-
22	l	r	-
23	r	g	-
24	g	l	-
25	r	g	-
26	l	g	-
28	r	r	+
30	l	r	-
31	l	l	+
32	r	r	+
33	l	l	+
34	l	r	-
35	r	l	-
36	l	g	-
38	l	r	-
42	l	r	-

absolute Häufigkeiten		
l	11	8
r	8	9
g	3	5
Summe:	22	22

absolute Häufigkeiten	
+	4
-	18
18%	

Mögliche Einflüsse der unterschiedlichen Untersuchungsräume

Raum A			
ID	Tester 1	Tester 2	Übereinst.
2	l	r	-
4	r	l	-
6	r	l	-
8	g	l	-
17	r	g	-
19	g	l	-
20	l	r	-
22	l	r	-
23	r	g	-
24	g	l	-
25	r	g	-
26	l	g	-
28	r	r	+
30	l	r	-
31	l	l	+
32	r	r	+
33	l	l	+
34	l	r	-
35	r	l	-
36	l	g	-
38	l	r	-
42	l	r	-

Raum B			
ID	Tester 1	Tester 2	Übereinst.
1	l	r	-
3	l	r	-
5	l	r	-
7	l	r	-
9	r	r	+
10	l	r	-
11	r	l	-
12	l	g	-
13	g	r	-
14	l	r	-
15	l	r	-
16	r	l	-
18	r	g	-
21	l	r	-
27	l	g	-
29	g	r	-
37	r	l	-
39	r	l	-
40	l	g	-
41	g	r	-
43	g	l	-

absolute Häufigkeiten			Summe
l	11	8	19
r	8	9	17
g	3	5	8

absolute Häufigkeiten			Summe
l	11	5	16
r	6	12	18
g	4	4	8

Häufigkeit der Untersuchungsergebnisse in Abhängigkeit der Tester

Interrater	Tester 1	Tester 2	Sicher*	Tester 1	Tester 2	Sicher**	Tester 1	Tester 2
ID	Ergebnis	Ergebnis	ID	Ergebnis	Ergebnis	ID	Ergebnis	Ergebnis
1	l	r	1	l	r	1	l	
2	l	r	2	l	r	2		r
3	l	r						
4	r	l						
5	l	r	5	l	r	5	l	
6	r	l	6	r	l	6	r	
7	l	r						
8	g	l						
9	r	r						
10	l	r						
11	r	l						
12	l	g						
13	g	r	13	g	r	13		r
14	l	r						
15	l	r						
16	r	l						
17	r	g						
18	r	g						
19	g	l						
20	l	r						
21	l	r	21	l	r	21	l	
22	l	r	22	l	r	22	l	
23	r	g						
24	g	l						
25	r	g						
26	l	g	26	l	g	26	l	
27	l	g	27	l	g	27	l	
28	r	r						
29	g	r	29	g	r	29		r
30	l	r	30	l	r	30	l	
31	l	l						
32	r	r						
33	l	l	33	l	l	33	l	l
34	l	r	34	l	r	34		r
35	r	l	35	r	l	35	r	
36	l	g	36	l	g	36	l	
37	r	l						
38	l	r	38	l	r	38		r
39	r	l						
40	l	g						
41	g	r						
42	l	r						
43	g	l						

absolute Häufigkeiten			absolute Häufigkeiten			absolute Häufigkeiten		
l	22	13	l	12	3	l	9	1
r	14	21	r	2	10	r	2	5
g	7	9	g	2	3	g	0	0
Summe:	43	43		16	16		11	6

12. Summary

The Test of the Abdominal Diaphragm Tension and its Reliability concerning the Right and Left Body (Interrater- Test Retest Study)

Context

The various appearances of a disease require an individualistic osteopathic approach to influence healing positively. The individual osteopathic examination is mainly based on intuition, which according to Sidler (2012) relies on experience, know-how and feeling. Liem (2012) stresses the importance of research to check the reliability of palpation and experience in order to support the daily work of an osteopath. Following Evidenced Practised Medicine (Walach, 2006) single case studies, qualitative analysis, repeated measurement designs, waiting list designs and reliability studies form useful research fields. The following parameters define challenges for reliability studies:

- Interaction and individuality of examiner and proband
- Treatment effect through examination
- Definition of equal palpation-depth and -location within the proband

Due to the described challenges of methodological examinations, the focus of a promising interrater test-retest study is on deliberate test design. A reliability study for a specific test may give evidence for a documented general diagnosis.

Abdominal diaphragm tension forms the specific test object for the described study.

The diaphragm forms a spanned shell between chest and abdomen and causes an inspiration movement by contraction. The movement increases abdominal pressure which is, according to Pickering (2002), transmitted to esophagus sphincter where an interaction of inspiration-, holding- and peristaltic muscles takes place. Following research of Williame and Finet (2007) tension and mobility of diaphragm and viscera influence each other. Kary (2009) describes a modified tension of body tissues because of traumata followed by compression forces. The diaphragm plays a decisive role to transmit tension either cranially or caudally to other body areas. Kary supports his anatomic and physiologic research with four descriptive case studies.

According to Helsmoortel (2002) and Meert (2007), the palpatory test of diaphragm may be conducted in a static, respiratory, dynamic, inhibitoric or provocative way in osteopathy. The reliability of these taught and described tests has not been verified yet.

Methodology

This study uses the dynamic test with induced translation into the frontal plane because it is both well described in literature (Helsmoortel, 2002, p.92; Meert, 2007, p.324; Münch, 2011, p.108; Paoletti, 2001, p.203) and presented in curriculum by Bernard Ligner, Dr. Erich Mayer-Fally and Christian Fossum. Furthermore the author of the study has an individual preference for it. Following Meert's Description (2007, p.324), the test of abdominal diaphragm tension is conducted in the following way (see figure 1):

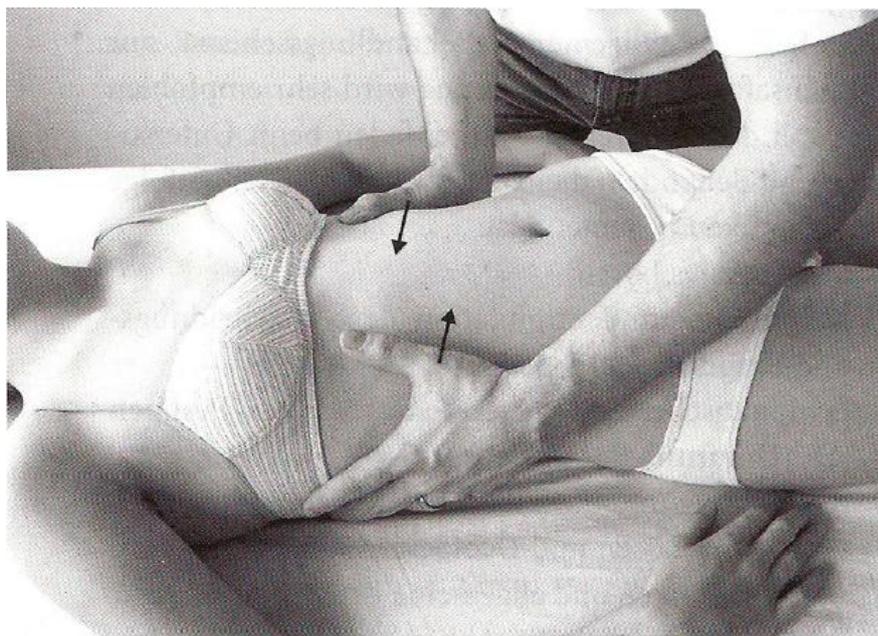


Fig 1: Test of abdominal diaphragm tension (Meert, 2007, p.324)

The examiner puts hands on both sides of the proband's costal arch coming from the lateral position. The beneath diaphragm is visualized. The proband's ribcage is passively and smoothly moved from one side to the other. The examiner judges the start of the movement, no maximum movement is performed. Furthermore the examiner decides whether tension is stronger or equal on any side. The side which moves better or allows better movement is identified.

Research Questions

Do two osteopaths who conduct an abdominal diaphragm test by translation come to similar results within a single proband? (interrater reliability)

Does one osteopath who repeats the abdominal diaphragm test on the same proband achieve a similar result (test-retest reliability)?

Do the test results coincide with the probands' chronic pain in the lumbar, thorakolumbar or thoracic area?

Study Design

The concept of methodological study was chosen to answer the research questions. In general reliability studies reveal a low level of trustworthiness in manual medicine (Haneline, 2009; Olivo, 2008; Patjin, 2004; Stovall, 2010). In order to raise this level, the International Federation for Manual/Musculoskeletal Medicine (FIMM) has proposed a handbook concerning methodological studies to reduce contrasting approaches (Patjin, 2004). The paper at hand follows the handbook in its structure.

Inclusion Criteria

Twenty persons who are both between the age of 18 and 75 and in a subjectively healthy condition were requested. Additionally the study required twenty test persons who featured pain in the lumbar, thoracolumbar or thoracic area for at least five days a week in the last three months (definition of chronicity according to Snider, 2008). Following Kuchera (2007) and Snider (2008) changes in the body tissue and movement become perceptible in cases of chronic pain in the thoracolumbar area. This observation may be the reason why differences in the tissue of the right and left body side are better to identify on patients with pain. Consequently the accordance in the test results is supposed to be higher than with healthy patients.

Exclusion Criteria

Due to contraindication in the examination technique, persons with urgent evolution in thorax and abdomen (pneumonia, pneumothorax, pulmonary embolism, fractures) were excluded. Furthermore women in their last trimester of pregnancy were not allowed to participate because of higher probability of recognition in the test-retest reliability check.

Subjects

A total of 43 probands aged 30-75 years (average = 56 years, spreading = 13,7 years; median = 60 years) took part in the study. 24 females and 19 males participated. 22 probands showed no signs of pain whereas 21 test persons of sample featured pain (nine probands in the left side, five in the right side and seven in both sides). Five probands showed pain in the thoracic area (Tx), two in the thoracolumbar passage (TLP) and fourteen in the lumbar area (Lx).

Recruitment

The probands were either chosen from the author's pool of patients or selected from relatives, friends and acquaintances. Colleagues and doctors did not react to phone calls/mailings and brought no additional test persons. The probands became informed about the course of the study through an information sheet. After the probands' data had been determined before, the test persons were grouped into patients with and without pain.

Randomization

The author decided about examination order of tester and probands by lot and was witnessed. Furthermore twenty test persons of the 43 probands were selected by lot to take part in the second test course dealing with test-retest examination (see figure 2). The examination order may give information about possible treatment effects in consequent examinations.

Testers

Two osteopaths both participated in the interrater and test-retest reliability. The chosen osteopaths have worked the same number of years (final exam at WSO 2006), reveal the same basic education (graduated physiotherapist at AKH Wien 1994-1998) and offer similar further training certificates (spiraldynamic, biodynamics by Jim Jealous). Furthermore both osteopaths work bimanually and write with their right hand.

Training

A test run with five probands and two testers helped to optimize the random sample, the course of the test, the management and the technical documentation. One month before the study the test instruction got sent by mail to the testers. The author recommended to apply the technique each work day on five patients and ask for their feedback afterwards. An instruction of the osteopaths took place on the examination day to harmonize force, speed, grip mode and way of test execution. Klein (2004, p.104) managed to ascertain an influence of tissue's viscoelastic characteristics by the named parameters.

Test Procedure

Figure 2 provides a schematic overview of the examination procedure. 43 probands (22 probands without pain and 21 probands with pain) took part in the interrater reliability test in a randomly selected order. This test was followed by the test-retest study in which twenty test persons (nine probands with pain and eleven probands without pain) participated.

Two separated examination rooms enabled a simultaneous test of two persons. Tester 1 was first to check in room A and second in room B, whereas tester 2 worked vice versa.

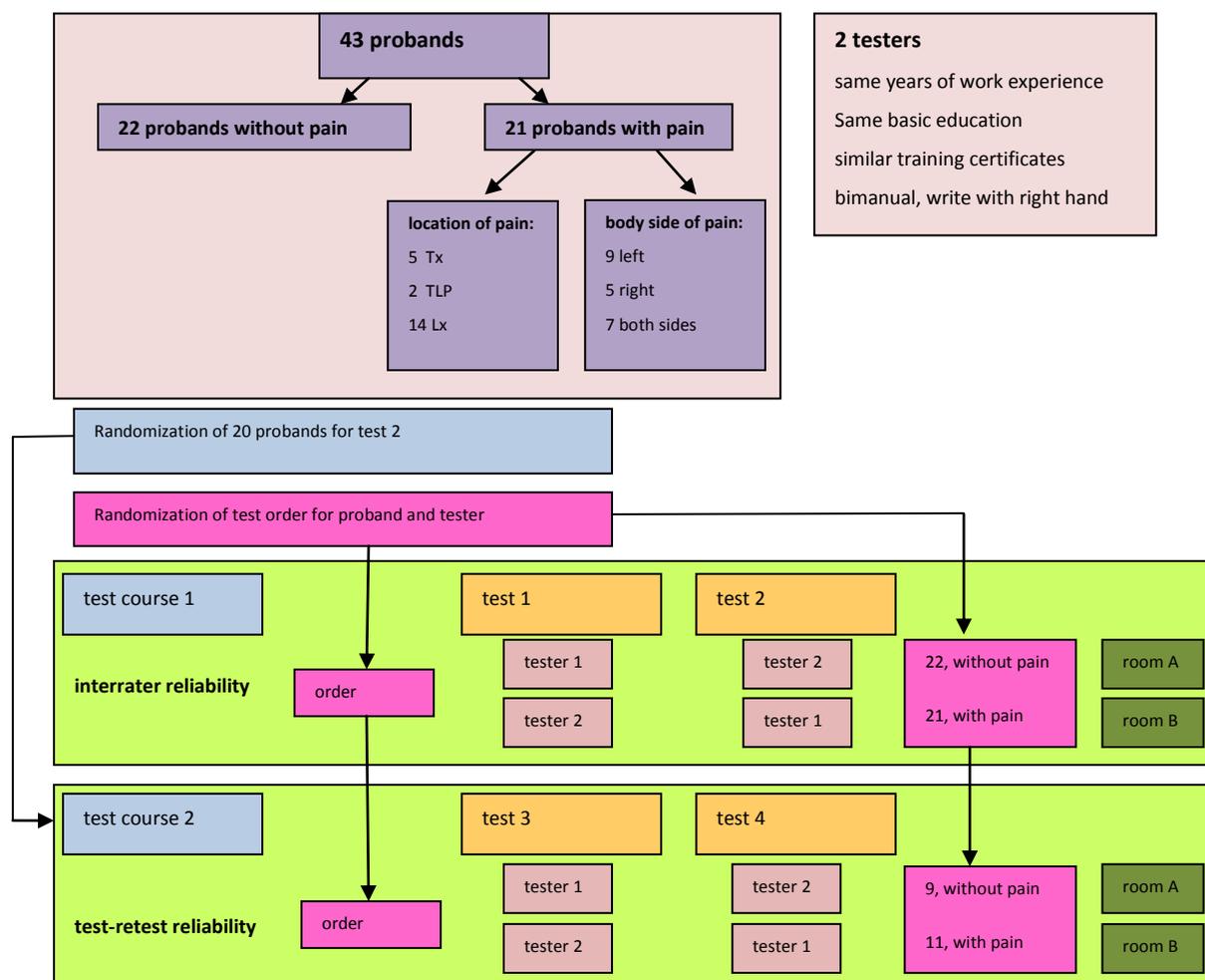


Fig 2: schematic overview of examination procedure

Before the test courses began, the examination table were adjusted to the size of the osteopaths. The assistants were briefed about their tasks. The author served as head of study and welcomed the arriving probands in the waiting room and delivered the questionnaires to be filled in.

The assigned numeration of the test person was written on a clearly visible sticker next to the clavicle. Two assistants accompanied the proband in casted order to room A and B. The third assistant blinded probands, informed the osteopath when to start and took notes of the results in room A whereas the fourth assistant did the same in room B.

The test persons were asked to lie down on the examination table as relaxed and calm as possible (laughing and holding one's breath considerably changes tension). When the osteopath was accompanied by the assistant into the examination room, the proband blindly lay on the examination table.

The osteopath positioned himself at lap height right to the proband and welcomed the test person with a soft touch on the proband's arm. The diaphragm tension test was done twice on each side. Before the test, hand signals between osteopath and assistant had been

agreed upon to make nonverbal communication possible. In case of sure palpation result, the assistants encircled the result. A study proves that coincidence in case of sure results is higher (Krause, 2008, p.8). After the examination course probands and testers had to fill in a feedback questionnaire.

Statistical Analysis

The handwritten results of the testers were transferred into a Microsoft® Excel®-Table. The interpretation of the results was computer assisted by the statistic software R 2.14.1. (R Development Core Team, 2011). The so called Kappa-Index (**κ -Index**) according to Cohen (1960) serves to judge reliability. Kappa index measures to what extent an observed accordance between two testers (interrater reliability)/ two tests (test-retest reliability) exceeds coincidence by chance. Value zero represents sheer chance, value +1 stands for total accordance. Negative values like in the presented study appear when the frequency of according results is lower than coincidental accordance and may reach the minimal value of -1. Kappa indices according to Cohen additionally serve to calculate possible treatment effects and accordance location of pain between test results.

χ^2 -tests (Chi-square-tests) serve to make further statistical examinations dealing with differing relative frequencies in case of accordance/ nonaccordance in test-retest reliabilities. Additionally χ^2 -tests provided results about the preferred bodyside tension of the two testers. Finally these tests are used to find out about existing influences and differing examination surroundings in room A and B.

Exact tests according to Fisher measure frequency distribution of accordance /nonaccordance depending on the sequence of examination. The significance level of $\alpha=0,05$ was chosen in all tests.

Results

Verification of Hypothesis

The first hypothesis has to be dismissed. The test method is fair reliable (interrater reliability $\kappa=-0,35$). The difference between the result of the two testers is significantly higher than their coincidental accordance can explain ($z=3,39$, $p=0,00070$). Therefore this result provides information about systematic differences between the two examiners.

Tester 1 achieves an accordance of fifteen out of twenty probands in the test-retest reliability (75%, $\kappa=0,57$ in the Landis and Koch scale, moderate reliability, $z=3,31$, $p=0,00095$, significant), Tester 2 reaches poor reliability 45% ($\kappa=-0,02$, equivalent to coincidental accordance). In terms of frequency levels accordance/ nonaccordance, the two tester do not differ in statistically significant way. If the difference in the accordance level between tester 1

and 2 was raised by the number 1 or 2, the result would be significant. Tester 1 confirms hypothesis 2, tester 2 does not.

In case of the specified bodyside of pain both testers do not reach an accordance level higher than the expected coincidental accordance (tester 1 with $\kappa=0,11$ $p=0,43$; tester 2 with $\kappa=-0,16$, $p=0,31$). Tester 1 achieves eight accordances whereas tester 2 manages five accordances. The calculated coincidental accordance is six accordances for tester 1 and seven accordances for tester 2. Therefore hypothesis 3 can not be confirmed.

Discussion

Discussion of Hypothesis 1

The fact that the two testers agree in their results less than coincidence provides significant consideration about systematic differences. The evaluation of the probands' feedback questionnaire makes differences between the two examiners concerning the interrater reliability visible. Secondary a survey about possible treatment effects was made. Finally potential influences of the examination rooms were statistically evaluated and the frequencies of the body side test results were compared between tester 1 and 2.

1. A majority of probands names differences in grip position (32 of 40) and pressure (30 of 40) in the questionnaire. Only five out of 40 test persons believe to be examined by the same osteopath. This described discrepancy between the two testers may have effected the differing interrater test results. However, it is statistically not significant.
2. No matter whether tester 1 or tester 2 was first to examine, the deviance was either significantly negative (tester 2 with $k=-0,41$, $p=0,0029$) or tended to result in $k=0$ (coincidental accordance) (tester 1 with $\kappa=-0,28$, $p=0,065$). No significant dependency concerning the order of examination and the frequency distribution of accordance/nonaccordance could be found (exact test following Fisher: $p=0,34$). Therefore a treatment effect is either not observable or occurs with both testers in similar scale.
3. An influence of differing environmental effects in the two examination rooms can be ruled out. The organisational study design ensured that the two testers examined both in room A and in room B. The χ^2 -test of frequencies in the two rooms concerning "left", "right" and "both" sides produced $p=0,89$ and confirms that no influence of examination environment on the tested diaphragm tension side existed.
4. It is safe to deduce individual preferences from the examination results. Tester 1 declares nine times "left" and only twice "right" whereas tester 2 considers his results five times sure on the right side and once on the left side. Following the exact test of Fisher the difference of the two testers' frequency to consider results "sure" is significant ($p=0,035$). Based on statistical analysis it can be assumed that tester 1

prefers to detect tension in the left body side whereas tester 2 rather concentrates on the right side of the body.

The total examination results of test 1 reflect these possibly preferred body orientations. Tester 1 tends to find tension in the left body side (22x vs 14x right side) while tester 2 locates most often tension in the right side (21x vs 13x left side). The result "both sides" is located seven times by tester 1 and nine times by tester 2.

Following Oschman (2002) different examiners may detect different symptoms on the same probands and interpret them in a differing way. According to Oschman every person possesses individual biomagnetic fields which interfere with each other in case of palpation. Therefore the contact between tester 1 and 2 with the same proband creates differing results and it makes a difference whether examiner 1 or 2 appear because of diverging magnetic fields. This produces differing examination results, changes emphasis in diagnosis and ends up in alternative therapeutic approaches.

Discussion of Hypothesis 2

Tester 1 shows a significantly higher level of reproducibility ($\kappa=0,57$, $p=0,00095$), but neither an accordance with the second tester (interrater reliability $\kappa=-0,35$, $p=0,0007$), nor with the expected body side of symptomatic patients ($\kappa=0,11$, $p=0,31$) is given.

Due to the study results it is recommended to design reliability studies with more than two testers. The reproducibility of examiner 1 is high whereas it is low in the case of examiner 2. This study cannot answer the question who of the testers is representative for osteopaths as a whole. It can only be assumed that the test is reproducible for some osteopaths and for others not. A higher number of testers would make statistical outliers more visible.

The author personally assumes that concentration on the test execution produced the low level of coincidence for tester 2. Maybe the direct instruction before the test ruled out the differing one month stage of practice before because of unequal routines. This procedure resulted in similar techniques but diverted from concentration on the skill of palpation.

If the examiners had to choose between "left" and "right" instead of "both", "left" and "right" then the result would have probably been more intuitive and correct in border cases.

Basically the comparison between the right and left diaphragm may be questioned because of the differing subjacent organs (liver on the right side, stomach on the left side) and their effect on environment. The depth and gravity of the liver and the hollow stomach may be interpreted differently by various examiners. Experience of Dr. Erich Mayer-Fally and Christian Fossum assess right diaphragm side with more tension in case of a healthy proband. This know-how correlates with the pattern of Zink which is 80% right in pelvis diaphragm, left in abdominal diaphragm, right in thoracic outlet and rotates left in tentorium

cerebelli (Pope, 2003). The increased tension on the right side causes diaphragm to take the easier way and rotate to the left side.

Discussion of Hypothesis 3

The relationship between pain pathology and tension may be considered a check for palpation's accuracy. Snider (2008) and Kuchera (2007) are only two of numerous authors who serve as references for the approach that pain constitutes the reason for a patient's treatment. Pain represents only one parameter next to tissue texture change, asymmetry, restriction and sensitivity for a dysfunction. Projections, compensations or mental components may easily be ignored when only pain and tension are focused. The author is aware of the fact that equalization of tension localization with pain localization represents a reduction.

The decidedness of a proband's purely subjective pain localization without additional medical evidence may also be questioned. Several discussions with test persons revealed a subjective uncertainty about the site of chronic pain (left, right or both sides).

Perhaps an additional prone palpation test of tissue texture change in the autochthonic dorsum muscles might have helped to confirm a subjective pain localization.

A preliminary inquiry of probands to check whether spinal blockades cause pain and restriction was considered before the test. However, a direct correlation between spinal blockades and changes in the diaphragm tension is not taken for granted. Besides an author's test on sitting probands would result in a further subjective effect and the test is still lacking validity. Therefore a preliminary inquiry was not conducted.

Due to uncertainty whether pain correlates in an ipsilateral way on the tissue level or rather serves as a compensation, projection or mental part it might have been an alternative to exclude pain patients from the study. Possibly the exclusion of the parameter pain would have meant sole concentration on symptomless test persons and resulted in clearer reliability tests.

Study Review

According to the testers the probands could not fulfill the requirement to lie in a relaxed way and breathe evenly. The examination was hampered by increased basic tension which was caused for example by holding one's breath or maximal breathing in and out of the probands. Maybe an assistant guided breathing in and out and a testing only in the phase of inspiration would have harmonized the starting position and reduced stress. The increased basic tension of the probands might have caused the testers' slight translational movement impulse to be too sensitive to be reproducible and significant.

A further aspect of review is the instruction of the osteopaths. The oral and written code of practice one month before the examination included a clearly defined description of the test

execution (the chapter "Methodology" was sent to them) but leaves a natural rest of individual interpretation to the testers. An individual way of test execution might have taken place in the testing month. The exercise in practice directly before the examination might have altered nuances of the testing exercise and caused a distraction from proper judgements. The informative value of the test might have decreased because of this change. Furthermore a differing test execution might have been accompanied with a different palpation depth of the testers. Following the probands' answers on the feedback questionnaire concerning a different finger pressure, the testers might have examined divergent tissues.

Following Liem (2006, p.151) a slight pressure gives information about tone, flexibility and turgor of the skin. Slightly more solid pressure reveals information about tone and turgor superficial muscle layers and solid pressure informs about deep muscle layers, fascial enclosures and the state of organs. In case of the present study one tester might have examined the tension status of the intercostal musculature whereas the second tester dealt with the status of the diaphragm.

Following the concept of tensegrity a differing palpation depth should still result in the same indication because the system shares tension in all layers. If intercostal musculature is raised in the left side, the left diaphragm should also reveal higher tension.

The extensively chosen area of pain from chest to lumbar vertebra could be regarded a shortcoming of the study. Possibly a narrowing of pathology in the thoracolumbar transition might have delivered more congruent results. However, the accomplishment to reach the needed number of test persons would have been made a lot more complicated.

Prospects

Divergent handedness of the two examiners would be an explanation to illuminate the individual preferences of the testers. However, both testers consider themselves bimanual and write with their right hand. Still it would be an interesting study to check whether left- and right handers generally produce individual preferences in a reliable test. In literature nothing could be found concerning this topic.

The tester's position during the examination could also be analysed. Do preferences occur, depending on handedness, when the examiner executes the test from the right or left side of the examination table?

The cancelling of blinded probands might have lower lowered the increased basic tension of the test persons. Probands were better prepared for the touch and consequently more relaxed in their basic tone which would facilitate palpation. Following Krause (2008) not only the tactile sense but all senses could be used for better results. Maybe the sole examination of interrater would suffice in the described case. A study with similar and comparable

parameters like the master thesis at hand could answer the question whether blinded test persons show more reliable results than examiners who use all their senses.

Medical relevant information for a diagnosis can only be filtered from a combination of several tests. Patjin (2004) warns of dependent results between selected tests in test combinations though. Snider (2008) managed to achieve coinciding test results dealing with "low back pain" results and two examiners. The results were won by judging four divergent parameters ("resistance to anterior springing, tenderness, tissue texture change, static rotational asymmetry"). Liem (2012) proposes a standardized diagnosis method (stringing together several tests) in order to compare ten osteopaths with each other. A complex validity test of several tests might verify or falsify a single test more appropriately than the traditional method.

For example: A diaphragm test with thumb palpation below the costal arch in dorsal position would serve as an interesting completion of the study at hand. This test considers diaphragm's shell form and sheds light on elasticity of the diaphragm. The test results of both tests could be compared with each other and in case of accordance both test were considered valid.

An important individual benefit of increased dealing with applied and literary aspects of a test lies in the required deepened accuracy. An operator of a test has to challenge parameters, has to know exactly why to use this test in this specific context and has to interpret an examination afterwards.

Intuition does not only rely on "gut instinct" but has a lot to do with experience and knowledge. A mutual reflection can definitely strengthen intuition (Sidler, 2012). A study of this kind may question individual treatment techniques, emblaze paradigms in producing diagnosis and contribute to the fact that intuition and knowledge together form a promising osteopathic approach. Retrospective case assessments, supervisions or writing of a masterthesis train knowledge, reveal personal paradigms and therefore help to a combination of conscious/unconscious thinking - called well - founded intuition.

Conclusion

The testers show a significant lower interrater reliability than sheer coincidental probabilities would have expected them to perform. Significant systematic differences exist between the two examiners. Tester 1 prefers the left body side, whereas tester 2 favors the right body side in the test result. Only tester 1 achieves a fair significant accordance in test-retest reliability.

This result can be interpreted in a way that the testers possibly react differently to ongoing tension in the probands' body sides. Examiners notice and interpret nuances differently.

Due to the fact that the testers show different results in their Kappa values at test-retest reliability, deduction and abstractions of the fact remain speculation.

A higher number of examiners would have enabled the author to generalize concerning the community of osteopaths and filter statistic outliers.

The testers observed an increased basic tone of the proband which might have hampered the palpatory test. Maybe a guided test exclusively during the phase of breathing in could have lowered the probands' basic tone. Additionally the cancelling of blinded probands might have decreased subconscious tension in the middle of the body.

In normal medical and clinical practise a preceding anamnesis and the patient oriented examination rhythm help a patient to relax. Consequently minor movement amplitudes aid to localize centres of tension in the body. A study which strings together several specific tests might allow better and more practical conclusions to establish diagnosis. However, the result would be a complex validity study rather than a typical reliability study.

Despite detailed research, a probationary examination and an instruction for the testers which aimed at reducing sources of error, the following proposals are made to optimize the test procedure:

An about one month lasting phase of familiarization with the test procedure should follow the testers' instruction. This instruction leaves grip position to the individual tester but provides well-founded background knowledge about diaphragm levels and especially the intended depth of palpation. A check of the test execution to achieve harmonization of palpation depths on the study author is considered useful. This check is again followed by a month-long phase of automotion. A probationary examination with statistical analysis can also disclose additional sources of error before a study test starts.

According to the author a sole concentration on healthy probands suffices. 30 test persons and six testers seem to provide a usable result.

A diaphragm test with thumb palpation below the costal arch is an interesting extension of the original study because it can both help to prepare probands' relaxation and serve as a reference value.

Still the test of abdominal diaphragm is both for the author and the test examiners of this study at hand a rewarding overview test to identify a patient's hypertone side. The subsequent treatment success advocates the combination of a diaphragm test with thumb palpation and a dynamic diaphragm test with transitional and rotationally induced movement. In osteopathy the reliability check executed like in the master thesis at hand has to be questioned. Individual freedom in the test execution and -interpretation is a promising approach when the complexity of human body and mind in patient and examiners is taken into account.

Acknowledgement

I would like to thank Dr. Erich Mayer-Fally for supervision of the master thesis, Dr. Gebhard Woisetschläger for statistical analysis, Robert Torlutter for help with the English translation and everyone, who took part in my study. In particular I thank my family for having supported me in manyfold ways.

References

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20: 37-46.

Finet G, Williame C. (2007): Deltadyn - Articular - Visceral. <http://www.deltadyn.be/visceral/ger/einleitung.htm> [28.03.2013].

Fritz, JM, Wainner RS. (2001): Examining Diagnostic Tests: An Evidence-based Perspective. *Physical Therapy* 81(9):1546-1564.

Haneline MT, 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.

Helsmoortel J. (2002): *Lehrbuch der visceralen Osteopathie*. Stuttgart. Thieme.

Kary DJ. (2009): The Transversus Thoracis Muscle in humans- diagnosis and treatment of associated pathology: an osteopathic perspective. *Journal of the American Osteopathic Association* 19(4):21-30.

Klein P, Sommerfeld P. (2004): *Biomechanik der menschlichen Gelenke–Grundlagen, Becken untere Extremität*. München. Elsevier.

Krause R. (2008): *Palpation – Wahrnehmung – Heilung*. Stuttgart. Sonntagverlag.

Kuchera ML. (2007): Applying osteopathic principles to formulate treatment for patient with chronic pain. *Journal of the American Osteopathic Association* 107(11)6:28-38

Liem T. (2006): *Morphodynamik in der Osteopathie*. Stuttgart. Hippokrates.

Liem T, Hilbrecht H, Schmidt T. (2012): *Osteopathie und Wissenschaft*. *Osteopathische Medizin* 13(1):11-17.

Meert GF. (2007): *Das venöse und lymphatische System aus osteopathischer Sicht*. München. Elsevier.

Münch G. (2011): Manuelle Stimmtherapie (MST), eine Therapie, die berührt. Idstein. Schulz-Kirchner Verlag.

Olivo SA, Macedo LG, Gadotti IA, Fuentes J, Stanton T, Magee DJ. (2008): Scales to assess the quality of randomized controlled trials: a systematic review. *Physical therapy* 88(2):156-175.

Oschman JL. (2002): Clinical aspects of biological fields: an introduction for health care professionals. *Journal of bodywork and movement therapies* 6(2):117-125.

Paoletti S. (2001): Faszien. Anatomie, Strukturen, Techniken, Spezielle Osteopathie. München. Elsevier.

Patijn J. (2004): Reproducibility and validity studies of Diagnostic Procedures in Manual/Musculoskeletal Medicine. Protocol formats. International Federation for Manual/Musculoskeletal Medicine. Maastricht.

Pickering M, Jones JFX. (2002): The diaphragm: two physiological muscles in one. *Journal of Anatomy* 201:305-312.

Pope RE. (2003): The common compensatory pattern: its origin and relationship to the postural model. *American Academy Osteopathic journal*. Verfügbar unter: google scholar: RE Pope - Am Acad Osteopath J, 2003 - virginiasmiles.com [Stand: 20.02.2013].

R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Sidler S (2012): Wie denken die Finger? Unbewusstes Denken als eine Grundlage der Intuition. *Osteopathische Medizin* 13(3):4-9.

Snider KT, Johnson JC, Snider EJ, Degenhardt BF. (2008): Increased incidence and severity of somatic dysfunction in subjects with chronic low back pain. *Journal of the American Osteopathic Association* 108(8):372-378.

Stovall BA, Kumar S. (2010): Reliability of Bony Anatomic Landmark Asymmetry Assessment in the Lumbopelvic Region: Application to Osteopathic Medical Education. Review. *Journal of the American Osteopathic Association* 110(11):667-674.

Walach H, Falkenberg T, Fonnebo V, Lewith G, Jonas WB. (2006): Circular instead hierarchical: methodological principles for the evaluation of complex interventions. Verfügbar unter: <http://www.biomedcentral.com/1471-2288/6/29> [Stand: 20.02.2013].