

Reliabilitätsstudie des Befunds nach LMO

Reliabilitätsstudie über die Befunderhebung der
Wirbelsäule nach der Methode der
Lien Mécanique Ostéopathique

Master Thesis zur Erlangung des Grades
„Master of Science“ in Osteopathie

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

niedergelegt
an der **Wiener Schule für Osteopathie**

von ***Claudia Hafen-Bardella***

Burgdorf, November 2009

Betreut von *Claudia Gamsjäger und Friederike Kaiser*
Übersetzt von *Jack Hafen*

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, sind als solche gekennzeichnet. Sämtliche Quellen und Hilfsmittel, die ich für die Arbeit genutzt habe, sind angegeben. Die Arbeit hat mit gleichem Inhalt noch keiner anderen Prüfungsbehörde vorgelegen.

Diese Arbeit stimmt mit der von der Gutachterin beurteilten Arbeit überein.

Burgdorf, 30.11.2009

A handwritten signature in black ink on a light blue background. The signature reads "C. Hafen" followed by a long horizontal stroke.

Claudia Hafen

Danksagung

Ein grosses Dankeschön gebührt meinem Ehemann Felix der mich in vielen Belangen unterstützt und begleitet hat. So hat er die statistische Arbeit übernommen und mit mir stundenlang das richtige Berechnungs-Modell evaluiert. Er war mein Support am PC und in allen administrativen Belangen.

Bedanken möchte ich mich bei meinen Osteopathen-Kollegen/-Kolleginnen Angelika van de Kraats, Regula Studer, Diane Mottier und Dominique Moreau. Ohne ihren Einsatz hätte die Studie nicht durchgeführt werden können.

Das Lektorat und die Übersetzung von Jack Hafen waren für mich sehr wertvoll.

Bei meiner Schwester Ester und deren Familie möchte ich mich für die kulinarische und administrative Unterstützung ganz herzlich bedanken.

Allen Probanden möchte ich ein grosses Lob aussprechen. Dank ihrem Einsatz und ihrer Zuverlässigkeit, konnte die Studie ohne Probleme durchgeführt werden.

Ohne die Hilfe und den Einsatz aller oben erwähnten Personen, hätte ich diese Arbeit nicht schreiben können.

Abstract (deutsch)

Schlüsselwörter: Reproduzierbarkeit, *LMO*-Befundmethode, *Lien Mécanique Ostéopathique*, *Wirbelsäulen-Befund*, *Partielle Totale Läsion*, *Dominante Läsion*, Intra- und Inter-Rater-Reliabilität.

Ziel: Die *Lien Mécanique Ostéopathique (LMO-Methode)* ist eine manuelle Untersuchungs- und Behandlungsmethode entwickelt von Chauffour P. und Prat E. Beim *Wirbelsäulen-Befund* werden die *Partielle Totale Läsion* (Summe aller positiv gefundenen Tests) und die *Dominante Läsion* (stärkster positiver Test) ermittelt.

Forschungsfrage: Sind die durch den *Wirbelsäulenbefund*, nach der *LMO-Methode*, ermittelte *Partiellen Totalen Läsion* und *Dominanten Läsion* reproduzierbar?

Methodologie: Methodologische Studie zur Ermittlung der Intra-Rater- und Inter-Rater-Reliabilität des *Wirbelsäulenbefundes* nach der *LMO-Methode*. Die Stichprobengrösse ist mit mindestens 30 Probanden festgelegt. Nach Ermitteln der Ein- und Ausschlusskriterien konnten 36 Probanden in die Studie eingeschlossen werden. Die Homogenität der Probanden konnte gewährleistet werden. Vier studienunabhängige Tester führen bei jedem Probanden den Befund je zweimal durch. Vor und während der Studie hatten die Tester keine Einsicht in die Akten der Probanden. Der Korrelationskoeffizient wurde mit der Formel von Holsti ermittelt. Zur Bestätigung der Intra-Rater-Reliabilität ist ein Koeffizient von grösser/gleich $r_H:0.75$ gefordert, für die Inter-Rater-Reliabilität $r_H: 0.6$.

Ergebnisse: Die Hypothese der Reproduzierbarkeit der Ermittlung der *Partiellen Totalen Läsion* konnte bestätigt werden. Der Intra-Rater-Reliabilitäts-Koeffizient ergab einen Mittelwert von $r_H: 0.794 (\pm 0.098)$. Der Inter-Rater-Reliabilität-Koeffizient ergab einen Mittelwert von $r_H: 0.677 (\pm 0.068)$. Für den Befund der *Dominanten Läsion* konnte die Hypothese in beiden Fällen nicht bestätigt werden. Der Mittelwert des Koeffizienten für die Intra-Rater-Reliabilität betrug $r_H:0.465 (\pm 0.501)$ und für die Inter-Rater-Reliabilität $r_H: 0.181 (\pm 0.191)$

Konklusion: Es gibt eine gute Reliabilität, sowohl für den Intra-Rater als auch den Inter-Rater, für die Ermittlung der *Partiellen Totalen Läsion* bei *Wirbelsäulenbefund* nach der *LMO-Methode*. Es gibt keine Reproduzierbarkeit für die Bestimmung der *Dominanten Läsion*.

Abstract (English)

Key words: Reproducibility, LMO finding method, *Lien Mécanique Ostéopathique*, spine findings, partial total lesion, dominant lesion, intra- and inter-rater reliability.

Objective: The *Lien Mécanique Ostéopathique* (LMO method) is a manual examination and treatment method developed by P. Chauffour and E. Prat. With the spine findings the partial total lesion (sum of all positive tests found) and the dominant lesion (maximum positive test) are determined.

Research Question: Are the partial total lesion and the dominant lesion reproducible when they are determined by the spine findings according to the LMO method?

Methodology: A methodological study of the intra-rater and the inter-rater-reliability of the spine findings according to the LMO method. The size of the sample has been fixed at 30 test persons at least. After applying the inclusion and the exclusion criteria 36 test persons could be included in the study. The homogeneity of the test persons could be guaranteed. Four testers, independent of this study, do the findings twice with every test person. The testers had no insight into the files of the test persons, neither before nor during the study. The correlation coefficient was determined with the formula of Holsti. For the confirmation of the intra-rater-reliability a coefficient of equal/greater than: r_H 0.75 is required, for the inter-rater-reliability r_H 0.6.

Results: The hypothesis of the reproducibility of the determination of the partial total lesion was confirmed. The Intra-rater-reliability-coefficient showed an average of r_H : 0.794 (± 0.098). The Inter-rater-reliability-coefficient showed an average of r_H : 0.677 (± 0.068). For the findings of the dominant lesion, the hypothesis in both cases could not be confirmed. The mean average of the coefficient for the Intra-rater-reliability was r_H : 0.465 (± 0.501) and for the Inter-rater-reliability r_H : 0.181 (± 0.191).

Conclusion: There is a good reliability, for both the Intra-rater as well as the Inter-rater, for the determination of the partial total lesion with the spine findings according to the LMO method. There is no reproducibility for the determination of the dominant lesion.

Inhaltsverzeichnis

Teil I

1. Einleitung	10
1.1. Problemstellung.....	10
1.2. Wissenschaftliche Herausforderung	12
1.3. Fragestellung.....	13
1.3.1. Fragestellung	13
1.4. Hypothesen	14
1.4.1. Hypothesen für die <i>Partielle Totale Läsion</i>	14
1.4.2. Hypothesen für die <i>Dominante Läsion</i>	15
1.5. Relevanz der Arbeit für die Allgemeinheit/Patienten.....	15
1.6. Relevanz der Arbeit für die Osteopathie.....	16
2. Grundlagen	17
2.1. Geschichte	18
2.2. <i>LMO</i> -Diagnose	19
2.3. Teilbefund der ersten <i>funktionellen Einheit</i> nach <i>LMO</i>	22
2.3.1. Diagnosemodell der <i>LMO-Methode</i> der Wirbelkörper.....	22
2.3.2. Ablauf einer <i>LMO</i> -Testung in der Studie.....	24
2.4. Reliabilität.....	24
Die Reliabilität wurde in der Vergangenheit bereits auf viele verschiedene Arten definiert.	24
2.4.1. Gütekriterien empirischer Forschung.....	24
2.4.2. Inter-Rater-Reliabilität	26
2.4.3. Intra-Rater-Reliabilität	26
2.4.4. Problematik der Reliabilitätsprüfung	26
3. Methodik	28
3.1. Studiendesign.....	28

3.2.	Anforderungen an die Tester.....	28
3.3.	Ein- und Ausschlusskriterien	28
3.4.	Datenerhebung.....	29
3.5.	Statistik.....	30
3.5.1.	Reliabilitätskoeffizient	31
3.5.2.	Holsti-Reliabilitätskoeffizienten (Rössler 2005: 190f).....	32
3.5.3.	Stichprobengröße, Normalverteilung, Signifikanz.....	34
3.6.	Vorstudie	35
3.7.	Auswirkungsanalyse: Failure Mode and Effects Analysis (FMEA)	36
3.8.	Durchführung der Studie	41
3.8.1.	Vorbereitende Arbeiten	41
3.8.2.	Rekrutierung, Ein- und Ausschluss	42
3.8.3.	Stichprobe.....	43
3.8.4.	Durchführung der Analyse, Ablauf	44
3.8.5.	Reaktion der Probanden auf den Untersuch.....	46
4.	Ergebnisse	47
4.1.	Ergebnisdarstellung.....	48
4.1.1.	Ergebnisse Intra-Rater	48
4.1.2.	Ergebnisse Inter-Rater	48
4.1.3.	Ergebnisse abhängig von Geschlecht und Symptom-Status	48
4.2.	Bestimmung der <i>Partiellen Totalen Läsion</i>	49
4.3.	Normalverteilung der Ergebnisse	51
4.4.	Bestimmung der <i>Dominanten Läsion</i>	52
4.5.	Einfluss Status – Symptomatisch – Asymptomatisch	54
4.6.	Einfluss Geschlecht.....	54
5.	Diskussion.....	55
5.1.	Reproduzierbarkeit der <i>Partiellen Totalen Läsion</i>	55

5.1.1.	Reproduzierbarkeit der Intra-Rater-Reliabilität.....	55
5.1.2.	Reproduzierbarkeit der <i>Inter-Rater-Reliabilität</i>	56
5.1.3.	Reproduzierbarkeit des Wirbelsäulenbefundes.....	57
5.1.4.	Handhabung und Standardisierung der Testung	57
5.1.5.	Einfluss der Sensibilität des Testers auf das Ergebnis	58
5.1.6.	Reproduzierbarkeit der <i>Dominante Läsion</i>	59
5.1.7.	Verändert die Testung der Facettengelenke die Rahmenbedingungen? 60	
5.2.	Wahl des Koeffizienten.....	61
5.3.	Vergleich mit anderen Arbeiten	62
5.3.1.	A review of intraexaminer and interexaminer reliability of static spinal palpation: a literature synthesis (2009)	65
5.3.2.	A systematic, critical review of manual palpation for identifying myofascial trigger points: evidence and clinical significance (2008)	65
5.3.3.	Reliability of physical examination for diagnosis of myofascial trigger points: a systematic review of the literature (2009).....	66
5.3.4.	Reliability of spinal palpation for diagnosis of back and neck pain: a systematic review of the literature (2004).....	66
5.4.	Vorschläge für anknüpfende Projekte.....	66
5.5.	Bedeutung der Ergebnisse für die Osteopathie und Ausblick.....	67
6.	Zusammenfassung	69
7.	Referenzen	71
7.1.	Abbildungen	71
7.2.	Tabellen.....	71
7.3.	Bibliografie.....	72
8.	Anhang.....	77
8.1.	Dokumente	77
8.1.1.	Probandeninformation.....	77

8.1.2. Schriftliche Einverständniserklärung	79
8.1.3. Stammbblatt der Probanden mit Ein- und Ausschlusskriterien	80
8.1.4. Befundblatt.....	81
8.2. Weitere Angaben zur Statistik	82
8.3. Einfluss Status – Symptomatisch – Asymptomatisch	83
8.4. Einfluss Geschlecht.....	85

Teil II

Summary (Short version of the Thesis in English).....	0-29
---	------

1. Einleitung

1.1. Problemstellung

In meiner Praxis arbeite ich seit 2005 ausschliesslich nach der von Paul Chauffour und Eric Prat (2003) entwickelten Methode der *Lien Mécanique Ostéopathique (LMO)* (siehe 2). Alle folgenden Begriffe in dieser Studie, die speziell die *LMO* betreffen, sind ein Zusammenschnitt (keine Zitate) aus den zwei Büchern, welche von den Begründern verfasst wurden (Chauffour, P und Prat, E; 2003 / Chauffour, P und Pratt, E und Michaud, J; 2009). Der Einfachheit halber für den Leser werde ich diese Begriffe in *Kursivschrift* kennzeichnen und im Kapitel 2.2. beschreiben und erklären.

Diese Methode (Chauffour, Pratt; 2003: 39ff) zeichnet sich durch einen standardisierten Befund (ca. 350 Tests) aus. Der Befund jedes Patienten wird nach dem gleichen Ablauf durchgeführt und protokolliert. Die Tests sind einfach und unkompliziert im Handling. Bei dieser systematischen Befunderhebung werden *acht funktionelle Einheiten* (siehe 2.2.) des Körpers nach *funktionellen Dysfunktionen* (Elastizitätsverlust) durchgetestet. Die *Dysfunktionen* werden mit einem *Provokationstest* (siehe 2.2) ermittelt. Dieser Test dient der Auffindung von erhöhter Gewebespannung in einer definierten Struktur im Gewebe. Die Empfindung dieser erhöhten Gewebespannung = *Dysfunktion* ist subjektiv geprägt. Anhand der gefundenen *Dysfunktionen* wird mittels *Inhibitionstest / Balancetest* (siehe 2.2) die *Dominante Dysfunktion* (siehe 2.2) jeder Einheit ermittelt. Da die Befunderhebung standardisiert und einfach ist, stellte ich mir die Frage, ob dies ausreicht, um reliable Ergebnisse bei einer Studie zu liefern. Bei meiner Literaturrecherche nach Reliabilitätsstudie im Bereich manueller Wirbelsäulenbefunde stellte sich heraus, dass genau die fehlende Standardisierung der Methoden als einen der Hauptgründe angegeben wurde, dass keine brauchbaren Resultate erzielt werden konnten (siehe 5.3) Das gleiche galt bei der Reproduzierbarkeitsstudie osteopathischer Tests von Rössel-Bretschneider et.al (2002).

Die Methode der *LMO* ist in der Osteopathie im Begriff, bekannt zu werden. So wurde sie in den vergangenen Jahren an verschiedenen Kongressen in Frankreich, Deutschland, Kanada und Russland vorgestellt. Fachartikel wurden in den

Fachzeitschriften Osteopathische Medizin, Elsevier (2007: 4-9) und in der Deutschen Zeitschrift für Osteopathie DO, Thieme (2008: 16-19) publiziert. Die Entwickler der Methode verfassten dazu drei Bücher. In verschiedenen Ländern und an verschiedenen Osteopathieschulen wird die *LMO-Methode* als Post-Graduates-Modul angeboten. Hand in Hand mit dem Steigen des Bekanntheitsgrades der Methode steigt auch die Nachfrage nach wissenschaftlichen Grundlagen zu dieser Methode.

Die Entwicklung und die Weiterentwicklung der Methode erfolgten bisher durch die Begründer selbst rein empirisch. Systematisch werden durch sie neue Tests gesucht und diese auf ihre Eignung hin untersucht. Nach einer Bewährungsphase werden sie anschliessend in den Untersuchungsgang aufgenommen und den Studenten dieser Methode weitervermittelt. Wissenschaftliche Arbeiten wurden bisher noch keine dokumentiert.

Anhand der Patientendokumentationen in meiner Praxis stellte ich fest, dass durch die Anwendung der *LMO* meine Untersuchungsergebnisse präziser wurden und ich hatte den Eindruck, diese reproduzieren zu können. Auch bemerkte ich anhand der statistischen Auswertung über die Häufigkeit der Interventionen, dass sich die Anzahl an Interventionen pro Patient verringerte. Diese Faktoren veranlassten mich, den osteopathischen Befund und die Behandlung nur noch nach dieser Methode durchzuführen. Mittlerweile gehöre ich dem Lehrkörper (Assistentin) der *LMO* an. Da ich mich täglich mit *LMO* auseinandersetze, sei es als Osteopathin oder als Assistentin, wurde der Wunsch immer grösser, mich um wissenschaftliche Grundlagen zu dieser Methode zu bemühen. So entschied ich mich, meine Masterthesis diesem Thema zu widmen.

Die Reliabilität (siehe 2.3.2) ist eines der Gütekriterien bei empirischer Forschung (Mayring 2005, 183ff) sie gibt an, ob das erzielte Resultat einer Methode reproduzierbar ist. Laut Mayring ist die Reliabilität einer Testmethode ein Kriterium für deren Qualität. Gibt es eine Reliabilität so kann darauf Aufbauend die Validität der Testmethode überprüft werden. Mit der Validität wird überprüft ob mit der Methode das gemessen wird was sie vorgibt zu messen (Mayring, 2005, 183). Die Erforschung der Reproduzierbarkeit der Methode ist also eine wichtige Voraussetzung, um darauf aufbauend weitere Studien machen zu können.

Der Nachweis der Reliabilität steht und fällt mit der Standardisierbarkeit der Versuchsbedingungen (Sommerfeld, 2006). Bei manuellen Techniken ist dies noch schwieriger und auch eines der Hauptprobleme bei bisher durchgeführten Studien. So erwies sich zum Beispiel bei der Reliabilitätsstudie von D. Liedtke (2004: 11-16), die fehlende Standardisierbarkeit des Tensionsbefundes des Thorax, als eines der Hauptprobleme für den Nachweis der Reliabilität.

Mittels einer Intra- und Inter-Rater-Reliabilitäts-Studie möchte ich ermitteln, ob die Standardisierung des Befundes hoch genug ist um bei den erzielten Befunden nach der *LMO-Methode* eine reproduzierbar zu sehen. Zu diesem Zweck nehmen an dieser Studie 4 Osteopathen und 36 Probanden teil. Jeder Osteopath untersucht jeden Probanden zwei Mal. Verglichen werden sowohl die Ergebnisse der verschiedenen Osteopathen untereinander als auch die Ergebnisse der zwei Untersuchungen des einzelnen Osteopathen.

1.2. Wissenschaftliche Herausforderung

Die Nachfrage nach komplementären Medizin-Methoden wurde in den letzten Jahren immer grösser. Entsprechend nahm die Anzahl der Patienten, die davon Gebrauch machen, deutlich zu (Carter, 2003). Dies warf bei Kritikern die Frage nach dem Nachweis der Wirksamkeit, Sicherheit und Wirtschaftlichkeit dieser Methoden auf.

Heute verfügt die klinische Forschung über ein breites Spektrum an Forschungsdesigns. Grimes und Schulz (2002) haben sie klassifiziert. Je nach klinischem Problem wird das geeignete Forschungsdesign gewählt. Soll aber die Wirksamkeit einer Methode überprüft werden, ist es laut Mayring (2005: 185) und Merten (1995:303) von Vorteil, wenn man vorab die Verlässlichkeit und Qualität der Methode, sprich deren Reliabilität, Validität und Objektivität ermittelt. Dies ist laut Mayring (2005:184-185) kein Muss und wird in quantitativer Forschung kaum praktiziert, aber es ist ein Vorteil, wenn man darauf verweisen kann. Die Reliabilität (Verlässlichkeit) einer Methode ist eine gute Voraussetzung, um danach deren Validität zu untersuchen. Die Reliabilität und die Validität sind unterschiedliche Gütekriterien mit unterschiedlichen Aussagen. Laut Mayring (2005:193f) macht es wenig Sinn, die Validität zu prüfen, wenn die Testmethode nicht vorher auf ihre Verlässlichkeit überprüft wurde. Umgekehrt braucht es aber keine Validität, um die Reliabilität zu prüfen.

Also entschied ich mich, als Erstes die Reliabilität der *LMO-Methode* zu prüfen. Da der Befund sehr umfassend ist, beschränkte ich mich bei dieser Studie auf die Erhebung eines Teilbefundes. Zu diesem Zwecke entschied ich mich, nur den Wirbelsäulen-Befund nach der Methode der *LMO*, als Teil aus der *ersten funktionellen Einheit*, isoliert zu untersuchen. Das Prüfen der Reliabilität beschränkt sich also ausschliesslich auf diesen Teilbefund.

1.3. Fragestellung

Es stellt sich die Frage, ob bei der Befunderhebung nach der *LMO-Methode* die Standardisierung hoch genug ist und so eine Reproduzierbarkeit des Befundes der Wirbelsäule festgestellt werden kann.

1.3.1. Fragestellung

Ist der Wirbelsäulen-Befund nach der Methode der *LMO* reproduzierbar?

Frage 1 (Intra-Rater-Reliabilität)

Gibt es eine Übereinstimmung der Ergebnisse, wenn ein Osteopath eine Befunderhebung der Wirbelsäule nach der Methode der *LMO* bei demselben Patienten mehrmals in einem gewissen Abstand und am selben Tag durchführt?

Frage 2 (Inter-Rater-Reliabilität)

Gibt es eine Übereinstimmung der Ergebnisse, wenn verschiedene Osteopathen eine Befunderhebung der Wirbelsäule nach der Methode der *LMO* bei denselben Patienten am selben Tag durchführen?

1.4. Hypothesen

Da der Befund nach der *LMO-Methode* zweiteilig ist, *Partielle Totale Läsion und Dominante Läsion* (siehe 2.3.), ergeben sich folgende Hypothesen:

1.4.1. Hypothesen für die *Partielle Totale Läsion*

Für die *Partielle Totale Läsion* der Intra-Rater-Reliabilität ergeben sich folgende Hypothesen:

Nullhypothese 1.1 (Intra-Rater):

Das Befundergebnis der Ermittlung der *Partiellen Totalen Läsion* der Wirbelsäule nach der *LMO-Methode* ist zufällig.

Alternativhypothese 1.1 (Intra-Rater):

Das Befundergebnis der Ermittlung der *Partiellen Totalen Läsion* der Wirbelsäule nach der *LMO-Methode* ist bei mehrmaliger Durchführung durch denselben Untersuchenden überwiegend gleich (reproduzierbar).

Für die *Partielle Totale Läsion* der Inter-Rater-Reliabilität ergeben sich folgende Hypothesen:

Nullhypothese 1.2. (Inter-Rater):

Das Befundergebnis der Ermittlung der *Partiellen Totalen Läsion* der Wirbelsäule nach der *LMO-Methode* ist abhängig vom Untersuchenden.

Alternativhypothese 1.2. (Inter-Rater):

Das Befundergebnis der Ermittlung der *Partiellen Totalen Läsion* der Wirbelsäule nach der *LMO-Methode* ist unabhängig vom Untersuchenden überwiegend gleich (reproduzierbar).

1.4.2. Hypothesen für die *Dominante Läsion*

Für die *Dominante Läsion* der Intra-Rater-Reliabilität ergeben sich folgende Hypothesen:

Nullhypothese 2.1. (Intra-Rater) :

Das Befundergebnis der Bestimmung der *Dominanten Läsion* der Wirbelsäule nach der *LMO-Methode* ist zufällig.

Alternativhypothese 2.1. (Intra-Rater) :

Das Befundergebnis der Bestimmung der *Dominanten Läsion* der Wirbelsäule nach der *LMO-Methode* ist bei mehrmaliger Durchführung durch denselben Untersuchenden überwiegend gleich (reproduzierbar).

Für die *Dominante Läsion* der Inter-Rater-Reliabilität ergeben sich folgende Hypothesen:

Nullhypothese 2.2. (Inter-Rater):

Das Befundergebnis der Bestimmung der *Dominanten Läsion* der Wirbelsäule nach der *LMO-Methode* ist abhängig vom Untersuchenden.

Alternativhypothese 2.2. (Inter-Rater):

Das Befundergebnis der Bestimmung der *Dominanten Läsion* der Wirbelsäule nach der *LMO-Methode* ist unabhängig vom Untersuchenden überwiegend gleich (reproduzierbar).

1.5. Relevanz der Arbeit für die Allgemeinheit/Patienten

Die Verlässlichkeit eines osteopathischen Befundes nach der *LMO-Methode* kann durch die Studie bestimmt werden. Meiner Meinung nach ist es wichtig, dass die Güte eines manuellen Befundes angemessen beurteilt werden kann. Der Osteopath muss wissen, wie gut sein momentaner Befund reproduzierbar ist, um die richtigen Schlüsse daraus zu ziehen. Dies dient dem Schutz der Patienten vor falscher Information aufgrund eines überbewerteten manuellen Befundes.

Vermeiden von unzweckmässiger Behandlung durch falschen Befund.

1.6. Relevanz der Arbeit für die Osteopathie

Durch die Reliabilität kann die Verlässlichkeit und Qualität (Mayring, 2005,183) eines osteopathischen Befundes der Wirbelsäule nach der *LMO-Methode* bestimmt werden. Falls die Nullhypothese stimmt, darf der Befund nicht überbewertet werden. Falls die Alternativ-Hypothese zutrifft, gewinnt der Befund nach dieser Methode an Qualität. Für eine gesicherte Aussage ist eine Studie jedoch nicht ausreichend.

Sollte sich erweisen, dass die Standardisierung der Methode hoch genug ist, um den Befund zu reproduzieren, so könnte die Studie der Weiterentwicklung eines zukunftsweisenden Befundmodells dienen. Vielleicht könnte dies einer der Schlüssel sein, um manuelle Befund-Methoden reproduzieren zu können. Sollte sich die Verlässlichkeit nicht nachweisen lassen, könnte an der Standardisierung weiter gearbeitet werden.

Falls die Verlässlichkeit der Befundmethode der Wirbelsäule nach *LMO* nachgewiesen werden kann, wäre dies ein Qualitätsmerkmal für diesen Teilbefund. Das Vermitteln und das Erlernen dieser Befund-Methode bekommen einen neuen Stellenwert. Der Unterricht der *LMO* baut auf einer standardisierten präzisen Instruktion auf. Ist der Befund reproduzierbar, so kann dies auch als Bestätigung der Genauigkeit des Unterrichts dieser Methode gewertet werden.

Die Ergebnisse ermöglichen Folgestudien: Laut Mayring (2005:185) macht es erst dann Sinn, die Validität einer Methode zu untersuchen, wenn zuvor deren Reliabilität untersucht wurde.

2. Grundlagen

Die *LMO* ist eine Methode zur Erstellung einer manuellen Diagnose und zur Durchführung einer auf dieser Diagnose beruhenden Behandlung des Körpers. Sie entspricht laut dem Begründer Paul Chauffour (Chauffour, Pratt, 2003: 13) dem osteopathischen Konzept:

- Die Struktur steht über der Funktion
- Suche die *Läsion*, behandle sie und lass das Weitere geschehen
- Die Arterie hat immer recht („The rule of the arteria is absolute“)

Die französische Bezeichnung „*Lien Mécanique Ostéopathique*“ bedeutet wörtlich „*mechanische osteopathische Vernetzung*“. Die Begründer der *LMO* benutzen für ihre Methode folgendes Modell als Grundlage:

Jedes Gewebe, welches embryologisch aus dem mesodermalen Blatt entstanden ist, bildet die Vernetzung in sich und unter allen übrigen Geweben (Somody-Neplaz, 2007: 5). Dieses Gewebe sind die Faszien und alles Bindegewebe (Chauffour, 2003: 15ff). Die durch diese Verbindung zugelassene Beweglichkeit nennen sie „*Osteofasziale Biodynamik*“. Eine Störung dieses Zusammenspiels wird von den Lehrern der *LMO-Methode* als *osteopathische Dysfunktion* bezeichnet. Zu dieser „*Osteofaszialen Biodynamik*“ nehmen Richter und Hebgén (2008: 25) in ihrem Buch wie folgt Stellung: Chauffour umschreibt in seinem Buch die Anheftung der Faszien am Skelett und deren Funktion übersichtlich. Er stellt die myofaszialen Ketten in Bezug auf die Körperbewegung dar und veranschaulicht sehr detailliert die biomechanischen Vorgänge in den einzelnen Bereichen der Wirbelsäule, des Thoraxes, der Extremitäten und des Kraniums. Interessant bei diesem Modell ist die myofasziale Erklärung zur Entstehung von *Dysfunktionen*.

Die Idee von Chauffour (2003: 39f) beruht darauf, dass *Dysfunktionen* durch entzündliche Prozesse im Bindegewebe und in den Faszien entstehen. Diese Entzündungen entstehen durch Überbelastung des Bindegewebes/Faszien, verursachen Narben und führen zu Elastizitätsverlust der Faszien und damit zu Bewegungsverlust. Nach diesem Elastizitätsverlust im Gewebe sucht er beim Patienten. Zum untersuchten Gewebe gehören: Bindegewebe, Knochen, Bänder, Sehnen, Muskeln, Herz, Niere, Blutgefäße, Faszien, Dura Mater und alle

Umhüllungen der Organe und Nerven. Dadurch dass der Fokus auf dieses Gewebe gerichtet ist, ist es im Sinne des *LMO-Modelles* möglich, neue bisher nicht bekannte *Dysfunktionen (Läsionen)* zu definieren, zu untersuchen und zu behandeln. Dazu gehören zum Beispiel *Dysfunktionen* des arteriellen Systems, der *intraossären Kraftlinien* (Chauffour, 2003: 109), der Unterhaut, des vegetativen Nervensystems und so weiter.

Russel PT, PhD (2004) schreibt in seiner Stellungnahme zum Buch von Chauffour 2003, dass die anatomischen Grundlagen, die Beschreibungen der faszialen Verbindungen und der daraus folgenden Funktionsstörungen ausgezeichnet sind. Er bemängelt aber zu Recht das Fehlen von Wirksamkeitsnachweis (dieser ist nur anekdotisch) und das Fehlen von unterstützender Literatur.

2.1. Geschichte

Die Entstehung der *LMO* (Sodynam-Neplaz, 2007, 4) geht auf das Jahr 1978 zurück. Die Methode zur osteopathischen Diagnose und Behandlung wurde von Paul Chauffour entworfen und entwickelt. Bei der Ausübung seines Berufs als Osteopath stiess Paul Chauffour immer wieder an seine Grenzen. Deshalb suchte er nach neuen Wegen. Gemeinsam mit Dr. Jean-Marie Guillot veröffentlichte er im Jahr 1984 ein erstes Buch über die Grundprinzipien der Methode mit dem Titel „*Le Lien Mécanique Ostéopathique*“. Kurz nach dem Erscheinen dieser Publikation gesellte sich Eric Prat zu Paul Chauffour. Im Laufe der vergangenen 20 Jahre haben Paul Chauffour und Eric Prat gemeinsam die Methode der *LMO* erweitert.

Ihr ständiges Anliegen war und ist es ihr Wissen weiterzugeben. Deshalb haben sie die Ergebnisse ihrer Arbeit in drei Büchern publiziert: Im Jahr 2002 in den Vereinigten Staaten „*The Mechanical Link*“, in Frankreich 2003 „*Le Lien Mécanique Ostéopathique*“ und 2009 „*Le Lien Mécanique Ostéopathique – Artères et système neuro-végétatif*“

Die *LMO-Methode* beansprucht für sich, ein eigenständiges, osteopathisches Diagnose- und Behandlungs-Modell zu sein.

2.2. LMO-Diagnose

Es folgen einige LMO-spezifische Begriffserklärungen (Chauffour, 2005: 39ff; Somodyn-Neplaz, 2007: 5-7):

1. Provokationstest

Es handelt sich um einen Test zur Überprüfung der Elastizität des Gewebes. Das verbindende Gewebe wird mittels gezieltem dreidimensionalem Zug oder Druck unter Spannung gesetzt (= *Provokation*). Resultiert daraus im Sinne der LMO keine Verformung, ist keine Elastizität vorhanden und der Test wird als positiv bewertet (*Dysfunktion* vorhanden). Nach diesem Verfahren wird der ganze Körper systematisch untersucht (ca. 350 Tests).

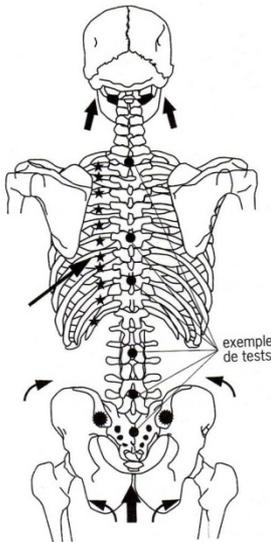


Abbildung 1 Provokationstest ganze Wirbelsäule
(Quelle: LMO Lehrkörper)



Abbildung 2 Provokationstest TH1
(Quelle: LMO Lehrkörper)

2. **Läsion = Dysfunktion:** positiver *Provokationstest*
3. **Totale Läsion:** Summe aller positiv gefundenen *Provokationstests* aller *acht funktionellen Einheiten*. Ziel jeder ganzheitlichen Behandlung nach der LMO-Methode ist es, die *Totale Läsion* zu finden und zu behandeln.
4. **Partielle Totale Läsion:** Summe aller positiv gefundenen *Provokationstests* bei Durchführung eines Teilbefundes. Der Teilbefund kann in einer oder in mehreren Einheiten durchgeführt werden. Wird bei „Notfallbehandlungen“ wie zum Beispiel unmittelbar nach einem Supinationstrauma angewendet.

5. **Inhibitionstest/Balancetest:** Bei der *LMO* nutzt man das Modell, dass die *Läsionen* untereinander gekoppelt sind. Daraus wurde der *Inhibitionstest* nach *LMO* entwickelt. Er dient der Hierarchisierung der durch den *Provokationstest* gefundenen *Läsionen*. Ziel dieser Hierarchisierung ist es, die Behandlungsdauer abzukürzen und so zeitsparend arbeiten zu können.

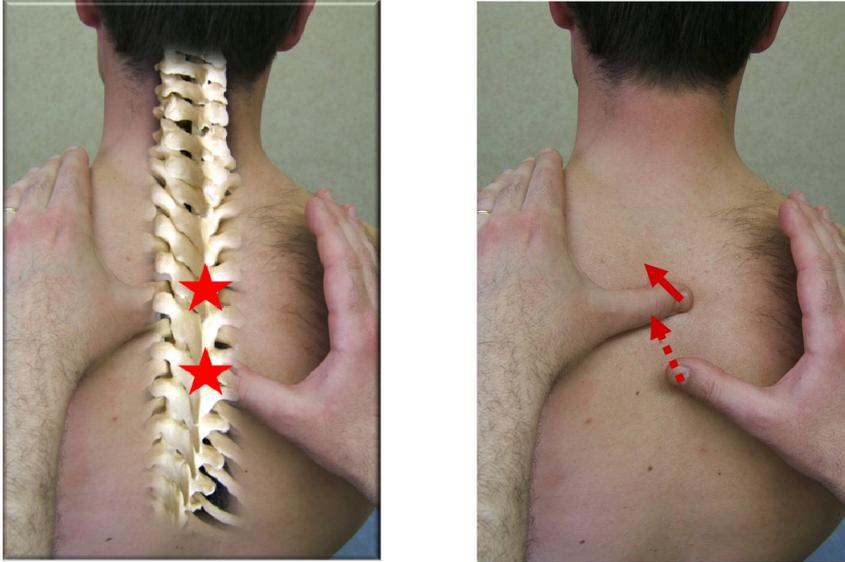


Abbildung 3 *Balancetest* BWS (Quelle: *LMO*-Lehrkörper)

6. **Primäre Läsion:** Stärkste *Läsion*, die durch den *Balancetest* bei der Hierarchisierung der Totalen *Läsion* (der positiven Tests aller *acht funktionellen Einheiten*) ermittelt wurde. Wird normalerweise bei jeder Behandlung bestimmt.
7. **Dominante Läsion:** Stärkste *Läsion*, die durch den *Balancetest* bei der Hierarchisierung der *Partiellen Totalen Läsion* (der positiven Tests der getesteten *funktionellen Einheiten*) bestimmt wird.
8. **Die acht Einheiten nach LMO:** Das Konzept der *LMO* hat den Anspruch, möglichst alle Strukturen im Körper testen zu wollen, das heißt die nach *LMO* benannte *Totale Läsion* zu finden. Da es ein komplexes Unterfangen ist, der Vielschichtigkeit der Anatomie mit einem Testverfahren gerecht zu werden, haben sich im Laufe der Zeit immer mehr Tests angesammelt. Diese mussten systematisiert werden. Das Anliegen an diese Systematisierung war, eine reproduzierbare Reihenfolge zu finden, die möglichst keinen Test vergessen lässt. Aus diesem Grunde wurde der Körper in die folgenden *acht funktionellen Einheiten* eingeteilt.

- 8.1. **Occipito-pelvine-Achse:** Wirbelsäule (WS), Becken, posteriorer Anteil der Rippen
 - 8.2. **Thorax:** Sternum, Clavicula, anteriorer Anteil der Rippen einschliesslich dem costochondralen Anteil
 - 8.3. **Extremitäten:** Gelenke, *intraossäre Läsionen*
 - 8.4. **Knochen:** *Kraftlinien* aller Röhrenknochen, Diastasen von Gelenken und Knorpelverbindungen
 - 8.5. **Cranium:** artikulär und intraossär
 - 8.6. **Viszera:** alle Organe
 - 8.7. **Gefässe:** Arterien, periphere Nerven, neuro-vegetatives Nervensystem
 - 8.8. **Haut:** am ganzen Körper
9. **LMO-Behandlung: „Recoil-Technik“:** Die *Recoil-Technik* ist eine durch Paul Chauffour entwickelte Technik, welche uns erlaubt, die *Primäre/Dominante Läsion* zu behandeln. Die *Recoil-Technik* ist eine direkte Technik. Der Beginn der Behandlung erfolgt bei der *primären/ dominanten* Struktur. Beendet wird die Behandlung, wenn kein *Provokationstest* mehr positiv ist.



Abbildung 4 *Recoil*, laterales Tibiaplateau nach posterior (Quelle: LMO-Lehrkörper)

Die Struktur, welche sich nicht verformen lässt, wird dabei in die eingeschränkte Richtung (dreidimensional im Raum) gedrückt und bis an ihre motorische Grenze unter Spannung gebracht. Durch einen leichten gezielten Impuls in die Richtung der Einschränkung wird dann diese Grenze durchbrochen. Anschliessend wird überprüft, ob das Gewebe seine natürliche Elastizität wieder erlangt hat, das heisst die Struktur verformbar ist. Dies wird mit dem ursprünglichen *Provokationstest* überprüft.

Zusammenfassend: Die *LMO-Diagnose* erfolgt in zwei Schritten und erfasst in der Regel immer alle *acht Einheiten*. Im ersten Schritt wird die *Totale Läsion* beziehungsweise die *Partielle Totale Läsion* mittels *Provokationstest* ermittelt und im zweiten Schritt werden die positiven *Provokationstests* mittels *Inhibitionstest* hierarchisiert, um so die *Primäre Läsion* respektive die *Dominante Läsion* bestimmen zu können (Somodyn-Neplaz, 2007: 7). Mit der Behandlung beginnt man bei der *Primären/Dominanten Läsion* und beendet sie sobald kein positiver *Provokationstest* mehr zu finden ist. Das heisst sobald die *Totale/Partielle Totale Läsion* aufgehoben ist.

2.3. Teilbefund der ersten funktionellen Einheit nach LMO

In meiner Studie möchte ich nur auf den Teil-Befund der *ersten funktionellen Einheit* eingehen. Diese Einheit umfasst Tests für die Wirbelkörper, die Rippen und das Becken. Ich beschränke mich auf die Testung (Somodyn-Neplaz, 2007: 6) aller Facettengelenke der Wirbelsäule mit dem *Provokationstest* nach der *LMO-Methode*. Dies ist ein verkürzter Befund, der aus Zeitmangel bei „Notfallbehandlungen“ eingesetzt werden kann. Nach erfolgtem Kurzbefund mit Behandlung in der Praxis am Patienten wird im Sinne der *LMO-Methode* so rasch als möglich ein ganzer Befund mit einer im Sinne der *LMO-Methode* ganzheitlichen Behandlung durchgeführt.

2.3.1. Diagnosemodell der LMO-Methode der Wirbelkörper

Wie oben beschrieben geht es bei der *LMO-Methode* darum, Elastizitätsverlust (Somodyn-Neplaz, 2007) zu finden. Beim Befund der Wirbelkörper will man die Elastizität gegenüber den überliegenden und unterliegenden Facettengelenken und Bandscheiben beurteilen. Hierzu wird als *Provokationstest* ein axialer Druck entsprechend den anatomischen Gegebenheiten im Verlauf des Dornfortsatzes ausgeübt (siehe Abbildung 5). Die Absicht dabei ist, sowohl das obere Facettengelenk als auch das untere Facettengelenk unter Spannung zu bringen.

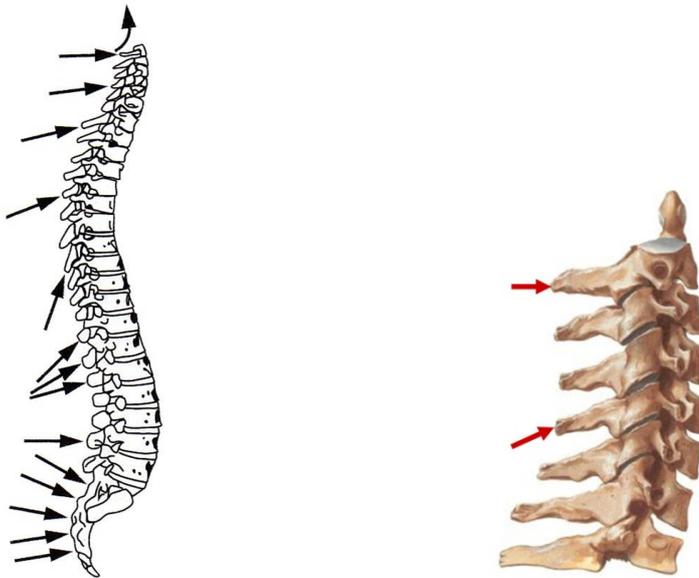


Abbildung 5 Axiale Druckrichtungen beim Provokationstest (Quelle: LMO-Lehrkörper)

Durch den axialen Druck wird der Wirbelkörper nach ventral gedrückt. Beurteilt wird, ob eine Elastizität spürbar ist oder nicht. Diese Beurteilung ist subjektiv und somit einer der kritischen Punkte bei einer Reliabilitätsprüfung. Am Patienten wird zuerst die *Partielle Totale Läsion* ermittelt. Anschliessend wird mittels *Inhibitionstest* die *Dominante Läsion* bestimmt.

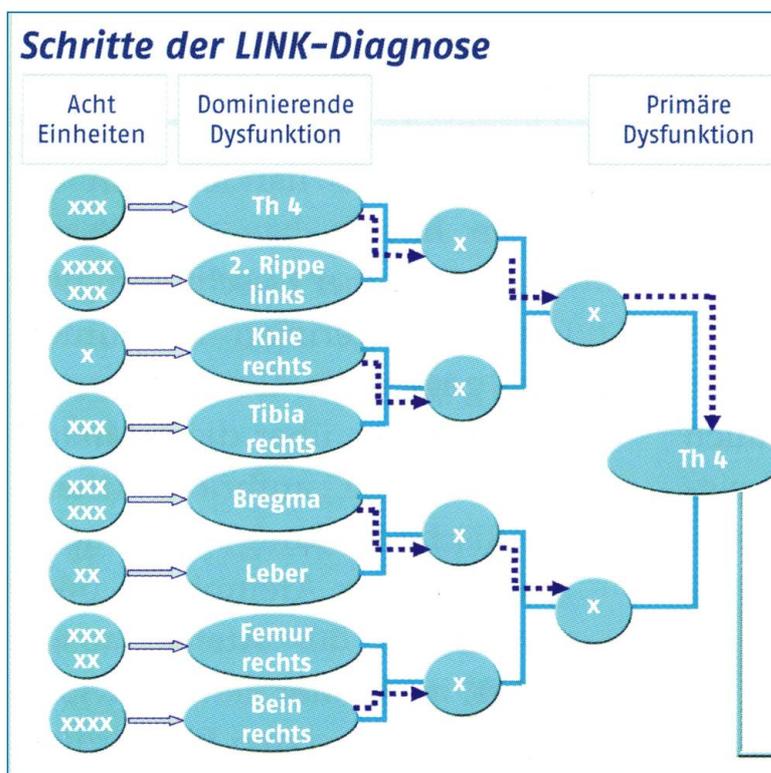


Abbildung 6 Ablauf der LMO-Diagnose (Quelle: LMO-Lehrkörper)

2.3.2. Ablauf einer LMO-Testung in der Studie

Zuerst wird mit dem *Provokationstest* (2.2.) die *Partielle Totale Läsion* (Chauffour, 2005) der Wirbelsäule ermittelt. Es werden pro Proband 25 Tests durchgeführt. In einem zweiten Schritt wird anhand der *Partiellen Totalen Läsion* mit dem *Inhibitionstest* die *Dominante Läsion* für die Wirbelsäule bestimmt.



Abbildung 7 Testung der Wirbelsäule (Quelle: LMO-Lehrkörper)

2.4. Reliabilität

Die Reliabilität wurde in der Vergangenheit bereits auf viele verschiedene Arten definiert.

2.4.1. Gütekriterien empirischer Forschung

Eines der Ziele in der empirischen Forschung (Mayring, 2005) ist es, dass erfasste Daten vergleichbar werden. Dies ermöglicht eine statistische Auswertung solcher Daten und lässt so die Möglichkeit offen, diese zu vergleichen und mögliche Abhängigkeiten zu erkennen. Aus diesem Grunde wurden Gütekriterien erarbeitet. Je besser eine Standardisierung von Untersuchungsinhalt, -ablauf und -situation ist, desto höher ist nach gängiger Auslegung die Bedeutung der entsprechenden Daten und Ergebnisse.

Nach Lienert (1989) unterscheidet man bei empirischen Untersuchungen Haupt- und Nebengütekriterien. Hauptkriterien sind die Objektivität, die Reliabilität und die Validität, Nebengütekriterien sind die Ökonomie (Wirtschaftlichkeit), Nützlichkeit, Normierung und Vergleichbarkeit von empirischen Untersuchungen. Weist eine Untersuchung diese Gütekriterien nicht auf, fehlen die wissenschaftlich überprüften Grundlagen und notwendigen Kontrolluntersuchungen.

Objektivität

(Lienert 1989): Objektivität ist das Ausmass, in dem ein Untersuchungsergebnis in Durchführung, Auswertung und Interpretation vom Untersuchungsleiter nicht beeinflusst werden kann, bzw. wenn mehrere zu übereinstimmenden Ergebnissen kommen. Unterschieden werden die Durchführungsobjektivität und die Interpretationsobjektivität. Bei der Durchführungsobjektivität geht es darum, dass das Untersuchungsergebnis vom Tester unbeeinflusst bleibt. Bei der Interpretationsobjektivität darf der Tester keine individuelle Deutung im Ergebnis einfließen lassen; es darf keinen Interpretationsspielraum geben.

Objektivität: "Die Objektivität eines Tests gibt an, in welchem Ausmass die Testergebnisse vom Testanwender unabhängig sind." (Bortz/Döring, 1995)

Reliabilität

Lienert & Raatz (1998: 9): Unter Reliabilität eines Tests versteht man den Grad der Genauigkeit, mit dem er ein bestimmtes Persönlichkeits- oder Verhaltensmerkmal misst, gleichgültig, ob er dieses Merkmal auch zu messen beansprucht. Der Reliabilitätskoeffizient gibt an, in welchem Masse unter gleichen Bedingungen gewonnene Messwerte über ein und denselben Probanden übereinstimmen, in welchem Masse also das Testergebnis reproduzierbar ist. Sie gibt eine Antwort auf die Frage: Wie genau misst der Test, was er misst?

Statista, 2009: Reliabilität (Zuverlässigkeit) ist ein wichtiges Kriterium für die Genauigkeit einer wissenschaftlichen Untersuchung. Damit ist gemeint, dass man unter gleichen Rahmenbedingungen zu den immer gleichen Ergebnissen kommen muss – es treten keine zufälligen Fehler auf. Wenn zum Beispiel ein Experiment wiederholt wird und dabei dieselben Ergebnisse erzielt werden wie bei der ersten Durchführung, spricht dies dafür, dass es reliabel ist.

Barkley, 2004: Reliabilität (auch Zuverlässigkeit): Unter Reliabilität versteht man den Grad der Genauigkeit, mit der eine Skala oder ein Test - bei verschiedenen Personen und wiederholt - ein bestimmtes Persönlichkeits- oder Verhaltensmerkmal misst (übrigens unabhängig davon, ob der Test dieses Merkmal auch zu messen beansprucht, dies ist ein Problem der Validität).

Zusammenfassend gilt: Die Reliabilität ist eines der drei Hauptgütekriterien und ist ein Mass für die Brauchbarkeit beziehungsweise die Zuverlässigkeit klinischer Tests oder Messmethoden.

Validität

Lienert (1989): Validität ist das wichtigste Testgütekriterium, denn es gibt den Grad der Genauigkeit an, mit dem eine Untersuchung das erfasst, was sie erfassen soll (z.B. Persönlichkeitsmerkmale oder Verhaltensweisen). Sie beantwortet die Frage: Misst der Test was er messen soll? Die Überprüfung der Gültigkeit wird mithilfe der Korrelation mit einem Aussenkriterium vorgenommen. Nach Mayring (2005: 185) macht die Überprüfung der Validität nur Sinn, wenn vorab die Reliabilität der Methode untersucht wurde.

2.4.2. Inter-Rater-Reliabilität

Mayring (2005: 185): Inter-Rater-Reliabilität bezeichnet in der empirischen Forschung (u.a. Psychologie, Soziologie, Epidemiologie,...) das Ausmass der Übereinstimmungen (= Konkordanzen) der Einschätzungsergebnisse bei unterschiedlichen Beobachtern (= "Rater" oder „Kodierer“). Ein Proband wird von zwei oder mehreren Osteopathen untersucht. Die Ergebnisse sollten für eine Inter-Rater-Reliabilität gleich sein. So kann angegeben werden, inwieweit die Ergebnisse vom einzelnen Osteopathen unabhängig sind. Die Reliabilität ist ein Mass für die Güte der Methode, die zur Messung einer bestimmten Variablen eingesetzt wird.

2.4.3. Intra-Rater-Reliabilität

Mayring (2005: 185): Die Intra-Rater-Reliabilität bezeichnet demgegenüber die Übereinstimmung der Einschätzungsergebnisse bei wiederholter Einschätzung durch denselben Beobachter. Hierdurch kann angegeben werden, inwieweit die Ergebnisse reproduzierbar sind und nicht von der momentanen Situation abhängen.

2.4.4. Problematik der Reliabilitätsprüfung

Um die Reliabilität zu steigern, müssen Zufallsfehler reduziert werden (Sommerfeld, 2006: 45). Zu den Verbesserungen gehören folgende Massnahmen:

- Standardisierung des Messablaufes
- Übung der Tester mit der Messmethode
- Automatisierter Messvorgang
- Messgerät optimieren

- Schaffen von konstanten Rahmenbedingungen
- Wiederholte Messungen

Es ist schwierig, Therapieeffekte bei wiederholter Testung auszuschliessen. Somit können sich durch die Testung die Rahmenbedingungen ändern.

Da es sich um manuelle Testungen handelt, ist die Messung immer subjektiv, respektive vom Empfinden des Testers geprägt.

Laut Früh (2007: 192) variiert die erreichbare Übereinstimmung mit folgenden Faktoren:

- Versiertheit, Sorgfalt und intellektuelle Fähigkeiten der Tester
- Eindeutigkeit und Vollständigkeit der angegebenen Indikatoren im Test
- Differenzierbarkeit des Kategoriensystems
- Hierarchische Struktur des Kategoriensystems
- Trennschärfe der Kategorien

3. Methodik

3.1. Studiendesign

Bei dieser Studie handelt es sich nach Portney & Watkins (2000) um eine methodologische Studie. Es wird eine Reliabilitätstestung durchgeführt. Untersucht werden die Intra-Rater-Reliabilität (siehe 2.4.3) und die Inter-Rater-Reliabilität (siehe 2.4.2) beim Befund der Wirbelsäule nach der Methode der *LMO*.

3.2. Anforderungen an die Tester

Um die in 2.4.4 beschriebenen Zufallsfehler möglichst klein zu halten, sollten die Tester mit der Methode der *LMO* vertraut sein. Sie müssen geübt sein, die *Partielle Totale Läsion* und die *Dominante Läsion* beim Wirbelsäulenbefund zu ermitteln. Deshalb ist die Anforderung an die Tester, dass sie die Methode seit mindestens drei Jahren täglich in ihren Praxen anwenden.

Bei den vier Testern handelt es sich um Kollegen, mit welchen sich die Studienleitung alle zwei Monate zwecks interner Weiterbildung im Bereich *LMO-Methode* trifft. Alle vier Tester wenden die Methode seit mindestens drei Jahren täglich und bei jedem Patienten an. Durch das häufige gemeinsame Üben sind die Abläufe der Tester gut standardisiert. Zwei Mitglieder dieser Erfahrungsgruppe gehören zum Lehrkörper von Chauffour und Pratt.

3.3. Ein- und Ausschlusskriterien

Bei der *LMO* geht es darum, Elastizitätsverluste des Gewebes zu diagnostizieren (Chauffour, und Prat, 2003). Chauffour und Pratt gehen davon aus, dass jeder Mensch solchen Bewegungsverlust hat, unabhängig vom Beschwerdebild. Bei dieser Studie beschränke ich mich auf Patienten mit unspezifischen nicht akuten Rückenschmerzen. Zum Zeitpunkt der Testung können die Beschwerden der Probanden symptomatisch oder asymptomatisch sein. Es dürfen keine radikulären Zeichen vorhanden sein.

Einschlusskriterien: Unspezifische Rückenschmerzen, symptomatisch oder asymptomatisch, mit und ohne Ausstrahlungen im Lenden- und Gesässbereich (nur Sensibilitätsstörungen).

Ausschlusskriterien: Radikuläre Symptomatik, radikuläre Zeichen wie verminderte Reflexe und Kraftverlust in Armen und Beinen, akute Rückenbeschwerden jeglicher Art, akute Entzündungen jeglicher Art, Fieber. Fehlende Daten bei der Befunderhebung der Wirbelsäule.

3.4. Datenerhebung

Mittels *Provokationstest* (siehe 2.2.) wird jeder Wirbelkörper der Wirbelsäule getestet. Dazu gehören: das Occiput (C0), die cervicalen Wirbelkörper 1-7 (C1-7), die thorakalen Wirbelkörper 1-12 (TH 1-12) und die lumbalen Wirbelkörper 1-5 (L1-5). Insgesamt werden bei jedem Probanden 25 Tests durchgeführt. Jeder *Provokationstest* muss kodiert werden und wird im Computer (PC) mit dem entsprechenden Vermerk positiv/negativ festgehalten. Bei jedem Probanden müssen pro Durchgang 25 Daten vorliegen. Aus den positiven Tests (*Partielle Totale Läsion*) wird die *Dominante Läsion* (siehe 2.3) ermittelt und ebenfalls im PC festgehalten. Der PC wird von nicht medizinisch geschultem Personal bedient. Der jeweilige Osteopath diktiert die positiven Tests dem Erfasser am PC. Mit diesem Vorgehen ist gewährleistet, dass der untersuchende Osteopath keine Akten- und keine PC-Einsicht hat. Er kennt während dem ganzen Ablauf der Testung nur den Namen des Probanden.

Die Erfassung erfolgt mit datenbankbasierten Formularen, welche anschliessend einfach in der Tabellenkalkulation ausgewertet und dargestellt werden können (siehe 8.2).

Vier Osteopathen, welche die Anforderungen gemäss 3.4. erfüllen, werden bei 36 Probanden je zwei Mal den Befund der ganzen Wirbelsäule nach der Methode der *LMO* durchführen. Die vier Tester dürfen während der Testung nicht zusammen kommunizieren (Früh, 2007:190) und bekommen keine Einsicht in die Akten der Probanden. Die Daten müssen vollständig vorliegen. Bei jedem Probanden müssen 25 Kodierungen vorliegen, die *Partielle Totale Läsion* muss ermittelt werden und die *Dominante Läsion* bestimmt werden. Zwischen dem ersten und dem zweiten Befund muss der Osteopath drei andere Untersuchungen durchführen. Damit ist gewährleistet, dass er sich beim zweiten Durchgang nicht mehr an den ersten Befund erinnert. Jeder Befund dauert maximal fünf Minuten.



Abbildung 8 *Provokationstest*
(Quelle: LMO-Lehrkörper)

Der *Provokationstest* erfolgt über axialen Druck im Verlauf der Dornfortsätze. Um eventuelle Nomenklaturfehler zu vermeiden werden die Dornfortsätze (DF) von C2-7, TH1-12, L1-5 beschriftet. Als Referenz für die Beschriftung der DF dienen die DF von C2, C7, TH1, TH12, L1 und L4 und L5. Diese Wirbel können dank Referenzpunkten leicht lokalisiert und definiert werden (siehe 3.7.). Von diesen Punkten aus werden die übrigen DF beschriftet. Die Beschriftung erfolgt bei allen Probanden durch die Studienleitung.

Durch die Studienleitung werden keine Befunde erhoben. Sie ist ausschliesslich für die Vorbereitung, Beschriftung und Durchführung der Studie zuständig (Sommerfeld, 2006: 86).

Die Anamnese mit den Ein- und Ausschlusskriterien wird anhand eines Protokolls (siehe 8.1.3) von einer Pflegefachfrau erhoben. Nötige Sicherheitstests werden von der Studienleitung, vor der Beschriftung der Wirbelkörper, durchgeführt.

3.5. Statistik

Rössler (2005: 190) und Früh (2007: 194-196) plädieren dafür, bei dieser Art von Studien die Berechnung der Reliabilitätskoeffizienten mit der Holsti-Formel durchzuführen. Sie stufen diese Methode als besonders geläufige Forschungspraxis, besonders für kleine Studien, ein. Auf Empfehlung des Statistikers wurde die Berechnung des Reliabilitätskoeffizienten r gemäss der Holsti-Formel vorgenommen und auszugsweise Berechnungen nach anderen Modellen zum Vergleich hinzugezogen (siehe 8.2)

3.5.1. Reliabilitätskoeffizient

Die Auswertung einer Reliabilitätstestung ist nur möglich, wenn ein Reliabilitätskoeffizient errechnet werden kann.

Definition: Reliabilitätskoeffizient = die Korrelation zwischen wiederholten Testungen, gibt den Anteil der wahren Varianz an der beobachteten Varianz an. (Merten, 1995: 304)

Laut Früh (2007: 196) eignet sich der Reliabilitätskoeffizient nach Holsti am besten für eine Reliabilitätsprüfung bei maximal 6 Testern und maximal 20-30 Kategorien. Allerdings ist es wichtig, dass man die Nachteile dieses Berechnungsverfahrens nach Holsti mit den entsprechenden Massnahmen beseitigt. Die Nachteile werden zusammenfassend von Hayes/Krippendorff (2007: 77-89) wie folgt beschrieben:

- Zahl der Variablenausprägungen: Die Höhe des Holsti-Koeffizienten wird beeinflusst von der Zahl der Ausprägungen, die eine Variable aufweist.
- Skalenniveau: Die Berechnung des Holsti-Koeffizienten ist nur für nominalskalierte Variable zulässig.
- Fehlende Werte: Die Berechnung des Holsti-Koeffizienten setzt voraus, dass die Werte aller Codierer komplett vorliegen.
- Präzision des Koeffizienten: Eine Berechnung von Konfidenzintervallen und damit die Angabe der Genauigkeit des Koeffizienten sind für den Holsti-Koeffizienten nicht möglich.
- Fehlende Korrektur für zufällige Übereinstimmungen: Die Berechnung des Holsti-Koeffizienten sieht keine Korrektur für zufällige Übereinstimmungen von Codierern beim Reliabilitätstest vor.

Diese Probleme können wie folgt ausgeschaltet werden:

Tabelle 1 Massnahmen für die Anwendung der Auswertung nach Holsti

Erforderliche Massnahme (Früh, 2007:190/194)	Getroffene Massnahme
Maximal sechs Tester	Vier Tester (siehe 3.2.)
Es dürfen maximal 20-30 Kategorien vorhanden sein	25 Messpunkte einer Kategorie, (siehe 3.5.3.)
Es braucht eine Nominalskalierung	siehe (3.4.).
Es dürfen keine Daten fehlen	Ausschlusskriterium siehe 3.3.).
Wegen der fehlenden Korrektur des Zufall-Faktors dürfen Koeffizienten um den Wert 0.5 nicht mehr akzeptiert werden	Bei mehreren Testern relativiert sich das Problem des fehlenden Zufall-Faktors (siehe 3.5.2.)
Die Tester dürfen nicht miteinander kommunizieren	Getrennte Räume für die Testung

3.5.2. Holsti-Reliabilitätskoeffizienten (Rössler 2005: 190f).

Der Holsti-Reliabilitätskoeffizient r_H beruht auf der Logik des Paarvergleichs.

Alle Messungen aller Tester (Codierer) bei allen Probanden werden also paarweise miteinander verglichen und die Anteile gegenüber der Gesamtheit bestimmt:

$$r_H = \frac{2 \times N_{\bar{u}}}{N_A + N_B}$$

r_H : Holsti's Reliabilitätskoeffizient

$N_{\bar{u}}$: Anzahl Übereinstimmungen zwischen Codierer A und B

N_A : Anzahl Codierungen (Einzeltests) von Codierer A

N_B : Anzahl Codierungen (Einzeltests) von Codierer B

Die Berechnung des Holsti-Koeffizienten erfolgt zweistufig: Zuerst berechnet man alle möglichen Paarvergleiche und bestimmt anschließend den Durchschnitt aller möglichen Paarvergleiche (Rössler 2005: 190)

Das Reliabilitätsprogramm PRAM (Neuendorf: 2009) erleichtert die Durchführung eines Reliabilitätstests wie dem von Holsti, denn es berechnet automatisch alle paarweisen Vergleiche und den entsprechenden Durchschnittswert. Wie aus Abbildung 9 ersichtlich ist, beträgt die prozentuale Übereinstimmung der Ergebnisse nach Holsti von Codierer 1 und Codierer 2 83.3%. Berechnet wurden alle Koeffizienten mit diesem Programm.

	A	B	C
1	Holsti's Coefficient of Reliability		
2			
3	Coder Pair		
4	1,2		0,833
5	1,3		0,667
6	2,3		0,667
7			
8	Average		0,722

Abbildung 9 Auszug einer Auswertung mit PRAM

Zusätzlich wurden die Resultate mit anderen Programmen verifiziert (siehe 8.2).

Für die Beurteilung der Güte der so errechneten Koeffizienten gibt es offenbar keine allgemein gültigen Richtwerte. Darüber herrscht in der Methodenliteratur offener Streit (Lomabard et. al:2008).

Es wird empfohlen, bei der Interpretation von Reliabilitätswerten die in der Methodenliteratur empfohlenen Schwellenwerte stets in Beziehung zum Schwierigkeitsgrad einer Variablen zu setzen. (Rössler 2005: 192). Dabei spielt die Anzahl Ausprägungen der Variablen eine wesentliche Rolle. Da bei dieser Studie allen Messungen nur zwei Zustände (positiv/negativ) zugeordnet werden, also dichotome Daten entstehen, liegt die Wahrscheinlichkeit einer zufälligen Übereinstimmung bereits bei 50 %. Würden zwei Tester bei einer grossen Anzahl Messungen also zufällig „raten“, wäre die Hälfte der Befunde wahrscheinlich übereinstimmend (Rössler 2005: 193). Also müssen die Werte, nach Rössler, bei der Anwendung von Holsti deutlich über 0.5 sein, da diese Formel keine Korrektur für zufällige Übereinstimmungen enthält. Für die Intra-Rater-Reliabilität (nur ein Tester) muss dies entsprechend berücksichtigt werden (siehe weiter unten).

Durch den Einsatz von mehr Testern wird dieser Effekt stark entschärft (Früh2007:195), da die zufällige Übereinstimmung der Ergebnisse eines Paares mit den Ergebnissen des nächsten Paares nur noch 50% von 50% beträgt. Bei vier Testern entstehen sechs Paarvergleiche mit je 50% Zufallswahrscheinlichkeit, folglich ergibt sich für die Übereinstimmung aller Paare nur noch eine Wahrscheinlichkeit von $50\%^{(6-1)} = 3.1\%$ (Rössler 2005: 193).

Somit ist es sinnvoll, für die Inter-Rater-Reliabilität, folgende Klassierung zur Darstellung der Ergebnisse anzuwenden, wie sie für andere Koeffizienten vorgeschlagen wird. (Cohens Kappa, Landis und Koch, 1977)

0.00 – 0.20 schlechte Übereinstimmung

0.21 – 0.40 vage Übereinstimmung

0.41 – 0.60 mässige Übereinstimmung

0.61 – 0.80 gute Übereinstimmung

0.81 – 1.00 sehr gute Übereinstimmung

Dabei ist wichtig zu wissen, dass diese Einstufung keinen Anspruch auf Absolutheit hat. Die entsprechenden Resultate können jedoch im Vergleich zu Ergebnissen vergleichbarer Studien durchaus eine hohe Aussagekraft haben.

Für die Intra-Rater-Reliabilität wird der Einfachheit halber dieselbe Klassierung verwendet. Dabei gilt zu berücksichtigen, dass diese Klassierung eine weitere Relativierung benötigt (siehe oben). Rössler (2005:192) fordert für formale Kategorien eine Übereinstimmung von nahezu 1.0 und für inhaltliche Kategorien Werte über 0.8. Bei der Beurteilung des manuellen Befundes von Positiv oder Negativ kann aber eher von einer inhaltlichen als von einer formalen Kategorie gesprochen werden. Also kann ein Wert um 0.8 durchaus als gut ausreichend eingestuft werden.

3.5.3. Stichprobengrösse, Normalverteilung, Signifikanz

Als Mindestwert für eine Reliabilitätsstudie werden in der Literatur 30 bis 50 Codierungen pro Kategorie angegeben. 200 bis 300 gelten als erstrebenswert (Früh, 1998: 189).

Mit der vom Statistiker empfohlenen Mächtigkeit der Stichprobe von 36 Probanden, 4 Testern, 2 Durchgängen und 25 Messpunkten ergibt sich eine grosse Anzahl an

Einzelmessungen (1800 Messungen für Intra-Rater-Test, 7200 Messungen für Inter-Rater-Test). Die 25 Messpunkte dürfen aufgrund der Gleichartigkeit näherungsweise als eine Kategorie angesehen werden. Damit ist die Wahrscheinlichkeit hoch, ein signifikantes Ergebnis für beide Tests zu erhalten.

Die nötige Mächtigkeit der Stichprobe ist abhängig von folgenden Faktoren (Hüsler & Zimmermann, 2006: 141ff):

- Homogenität der Stichprobe
- Anzahl Messungen pro Proband
- Anzahl Tester
- Zufälligkeit der Stichprobe (massgebend für die Möglichkeit der Verallgemeinerung der Ergebnisse)

Es soll auf eine gute Durchmischung der Probanden nach Geschlecht und Alter geachtet werden. Ebenso sollte der Anteil an symptomatischem und asymptomatischem Status ausgeglichen sein.

Bei den Messresultaten kann unter diesen Voraussetzungen eine Normalverteilung erwartet werden. Dies zeigt sich auch bei der Auswertung (siehe 4.3). Demzufolge wird die Standardabweichung s (SDEV = Standard deviation) für jeden Mittelwert \bar{x} (AVG = average: arithmetisches Mittel) ermittelt. In der Bandbreite von $AVG \pm SDEV$ befinden sich etwa $2/3$ der Daten; im Bereich $AVG \pm 2SDEV$ etwa 95% aller Messwerte (Hüsler&Zimmermann, 2006:38).

3.6. Vorstudie

Um eventuellen Problemen bei der Durchführung der Studie vorzubeugen, führte ich im Februar eine Vorstudie durch. Dabei wurde bei 6 Probanden von 3 Osteopathen je 1 Mal der Befund der Wirbelsäule nach der *LMO-Methode* durchgeführt.

Probleme ergaben sich bei der Beschriftung der Wirbelkörper, der Haltung der Probanden, der Verschiebung der Beschriftung beim Bewegen des Probanden und so weiter. Dies und weitere vermutete Fehlermöglichkeiten veranlassten mich, eine Fehlermöglichkeits- und Einflussanalyse durchzuführen.

3.7. Auswirkungsanalyse: Failure Mode and Effects Analysis (FMEA)

(deutsch: Fehlermöglichkeits- und Einflussanalyse)

Bei einer Reliabilitätsstudie in der Form wie sie hier durchgeführt wird ist es wichtig, dass die Reproduzierbarkeit unter ähnlichen bzw. gleichen Bedingungen geprüft wird (Kool & de Bie: 2001). Also muss darauf geachtet werden, dass möglichst konstante Rahmenbedingungen während der Testung geschaffen werden.

Wenn mit Menschen gearbeitet wird, ist es nicht möglich, für die Testung Laborbedingungen zu schaffen. Umso wichtiger ist es meiner Meinung nach, Zeit darauf zu verwenden, im Vorfeld nach möglichen Fehlerquellen und deren Auswirkungen zu suchen, sie zu erkennen und diese falls möglich zu beheben oder zu minimieren. Also müssen alle erkennbaren Einflussfaktoren mit Massnahmen soweit möglich wie günstig beeinflusst werden.

In der durchgeführten Vorstudie (3.6) kamen einige Probleme zum Vorschein. Um die Übersicht über diese Probleme und die einzuhaltenden Kriterien bei der Anwendung der Holsti-Formel (siehe 3.5.2.) zu behalten, entschied ich mich, eine Fehlermöglichkeits- und Einflussanalyse = FMEA (DIN EN 60812:2006) durchzuführen. Solche Analysen werden zum Beispiel in der Autoindustrie systematisch durchgeführt (ISO/TS 16949:2008) und dienen dem Auffinden versteckter Probleme und helfen Massnahmen zu definieren für deren Lösung.

Im Folgenden wird diese Fehleranalyse zusammenfassend in Tabellenform dargestellt:

Tabelle 2 Zusammenstellung FMEA

Fehler / Effekt	Ursache	Massnahme
Unterschiedliche Ausführung des Tests	Unterschiedliches Handling	Die Methode muss von den Testern mindestens seit drei Jahren praktiziert werden
	Unterschiedlicher Druck beim <i>Provokationstest</i>	
	Zu wenig Übung	

Fehler / Effekt	Ursache	Massnahme
Nicht einheitliche Benennung und Lokalisation der Wirbelkörper	Unterschiedliche Referenzpunkte, manuelle Festlegung (Interpretation) des Wirbelkörpers	Vorgängige Beschriftung der Dornfortsätze durch eine unabhängige Person, Instruktion, wie die Beschriftung anzuwenden ist
	Durch Positionsänderung des Probanden verschiebt sich die Beschriftung	Position der Probanden definieren und instruieren Probanden bleiben für jede Testung am selben Ort, der Tester wechselt den Raum, der Tester überprüft die Haltung vor der Testung
	Falsche Anwendung der Beschriftung (Palpationsfehler, Annäherung an Dornfortsatz)	Instruktion der Tester, wie die Beschriftung zu verwenden ist Genaue Definition, an welcher Stelle des Dornfortsatzes der Kleber angebracht wird
Beeinflussung der Tester	Mithören was der Vorgänger diktiert	Getrennte Räumlichkeiten für die Testung
	Einsicht in die Probandenakten	Keine Einsicht in die Akten

Fehler / Effekt	Ursache	Massnahme
Auswirkung der Testung auf den Bewegungsapparat	Mobilisation der Wirbelsäule durch die Ventralisation	Pro Wirbelkörper nur eine Testung pro Durchgang
	Lange Testzeit	
	Vegetative Reaktionen durch die Ventralisation	
Falsch erfasste Daten	Verwechslung bei der Eingabe am PC	Rückmeldung der erfassten Daten
	Missverständnis zwischen Tester und erfassender Person	Wiederholung des zu erfassenden Wirbels

Bei der Ausarbeitung dieser Analyse unterteilte ich die Auswirkungsfaktoren in zwei Gruppen: in beeinflussbare Faktoren und in nur bedingt beeinflussbare Faktoren.

Beeinflussbare Faktoren: Die Beschriftung der Wirbelkörper mittels Klebeetikette ist ein wichtiges Hilfsmittel, um Nomenklaturprobleme zu umgehen. Bei dieser Beschriftung geht es ausschliesslich darum, dass jeder Tester vom selben Wirbel spricht, unabhängig wie er heisst. Die Beschriftung erfolgt nach bestimmten Kriterien. Gut lokalisierbar sind die Dornfortsätze von C2, C7/TH1, TH12/L1 und L4. Diese werden zuerst als Referenzpunkte markiert und von da aus erfolgt die Beschriftung der übrigen Dornfortsätze. Der Dornfortsatz von C2 ist der erste, auf den man unterhalb vom Occiput stösst. Der Dornfortsatz von C7 ist gegenüber dem Dornfortsatz von TH1, der von der ersten Rippe zurückgehalten wird, bei einer maximalen Reklination der Halswirbelsäule (HWS) nicht mehr spürbar. Platziert der Osteopath einen Finger auf den Dornfortsatz von C7 und einen Finger auf den Dornfortsatz von TH1, so verliert der Finger von C7 am Bewegungsende den Kontakt zum Dornfortsatz von C7. Falls die Anatomie der Probanden der Norm entspricht, können so die Dornfortsätze von TH1 und C7 lokalisiert werden. Beim Dornfortsatz von TH12 kann man sich an der 12. Rippe orientieren und zusätzlich an der

schmalen bis punktförmigen Form des Dornfortsatzes (Sobotta, 2007: 260). Der Dornfortsatz von L1 ist hingegen breit und massig. Als Referenz für den Dornfortsatz L4 dient die Höhe des Beckenkamms. Da das Auffinden der Referenzpunkte durch Palpieren erfolgt, ist es wichtig, dass nur eine Person die Beschriftung macht. Die Tester werden am Skelett von dieser Person instruiert, wie die Beschriftung in Bezug auf den Dornfortsatz anzuwenden ist. Die Beschriftung wurde so gewählt, dass von dieser Stelle aus der *Provokationstest* gemacht werden kann. Ob der beschriftete Wirbel wirklich der Beschriftung entspricht, spielt bei dieser Arbeit keine Rolle. Die Beschriftung muss mit Klebeetiketten erfolgen, da sich beim Schreiben mit einem Stift die Haut verschiebt und so der Strich nicht an der richtigen Stelle angebracht werden kann. Auch der Druck, den man ausüben muss damit die Haut beschriftet ist, hinterlässt Spuren.



Abbildung 10 Beschriftung und Haltung der Probanden

Eine weitere Fehlerquelle ist das Verschieben der Beschriftung durch das Bewegen des Probanden. Die Testung erfolgt im Zeitraum von 40 Minuten. Es ist nicht möglich, in diesem Zeitraum still zu sitzen. Um auch diese Fehlerquelle klein zu halten, bleibt der Proband während der ganzen Testung am selben Ort und die Ausgangslage für die Testung wird so gewählt (siehe 3.7.), dass sie für den

Probanden komfortabel ist und er ohne Probleme 40 Minuten diese Haltung einnehmen kann. Zusätzlich ist der Tester aufgefordert, vor der Testung die Haltung des Probanden zu überprüfen und gegebenenfalls zu korrigieren. In der Zeit, in der die Tester den Raum wechseln, kann sich der Proband ohne Einfluss auf die Beschriftung leicht bewegen, aber während der Testung ist die vorgegebene Haltung einzunehmen.

Damit die Tester keine Akteneinsicht haben, werden die Angaben vom Hilfspersonal am PC eingegeben.

Damit keine Beeinflussung der Tester durch sehen und hören stattfinden kann, werden die Probanden in verschiedenen Räumen platziert.

Auch die Datenerfassung birgt einige Fehlerquellen. Die Daten werden diktiert und von Helfern am PC eingegeben. Die Helfer am PC wiederholen vor der Eingabe den diktierten Wirbel und wiederholen den markierten Wirbel nach der Eingabe erneut. Der Tester hat so die Möglichkeit, die Eingabe zu überprüfen. Damit die anderen Tester nicht mithören können, werden die Probanden in getrennten Räumen platziert.

Bedingt beeinflussbare Faktoren: Die Idee des *Provokationstests* beruht darauf, den Wirbelkörper nach ventral zu schieben und so die Elastizität der Facettengelenke zu testen. Falls der Test wirklich eine Bewegung impliziert (der Test valide wäre), stellt sich die Frage, ob es zu einer Mobilisation dieser Facettengelenke kommt. Falls dies der Fall wäre, stellt sich die nächste Frage, ob durch die Testung bereits eine Behandlung erfolgt und so der nächste Tester andere Bedingungen vorfindet. Immerhin werden die Facettengelenke bei nur einer Testung pro Durchgang acht Mal getestet. Laut Chauffour (2005) reicht die Testung bei einer *Dysfunktion* nicht aus, um eine Mobilisation in den Facettengelenken zu bewirken, sie bleibt bestehen. Diese Aussage ist nicht belegt. Eine weitere Frage stellt sich bezüglich des sympathischen Grenzstranges, der durch die Ventralisation stimuliert werden könnte, da er anatomisch direkt vor den Facettengelenken der Brustwirbelsäule (Netter) liegt. In der klassischen Osteopathie ist die Pumptechnik (Chaitow, 2002: 220), bei der die Brustwirbelkörper rhythmisch nach ventral gedrückt werden, eine gängige Technik zur Stimulation des neurovegetativen sympathischen Grenzstranges. (Die Validität dieser Technik ist nicht belegt.) Sind durch die Testung

vegetative Reaktionen zu erwarten? Falls solche auftreten, verändern diese die Rahmenbedingungen? Um all diese nicht berechenbaren Einflüsse etwas zu minimieren, sind die Tester aufgefordert, den Test nur ein Mal pro Wirbel und Durchgang durchzuführen. Da alle vier Tester diese Testung seit mindestens drei Jahren täglich anwenden, ist davon auszugehen, dass dies möglich sein sollte.

Nicht beeinflussbare Faktoren: Sowohl die Individualität der Probanden als auch das subjektive Empfinden der Tester ist gegeben.

3.8. Durchführung der Studie

Es mussten einige Dinge erarbeitet, geplant und koordiniert werden, damit der Ablauf reibungslos über die Bühne gehen konnte.

3.8.1. Vorbereitende Arbeiten

Folgende Dokumente mussten verfasst werden (Dokumente siehe Anhang):

- Probandeninformation
- Schriftliche Einverständniserklärung
- Erstellung eines Zeitplanes über den Ablauf der Befunderhebung und das Aufbieten der Probanden. 36 Probanden wurden an einem Tag acht Mal untersucht.
- Stammbblatt der Probanden mit Ein- und Ausschlusskriterien
- Befundblatt
- Datenbank zur Erfassung der Daten und Berechnungshilfen erstellen/beschaffen, damit der Reliabilitätskoeffizient errechnet werden kann.

3.8.2. Rekrutierung, Ein- und Ausschluss

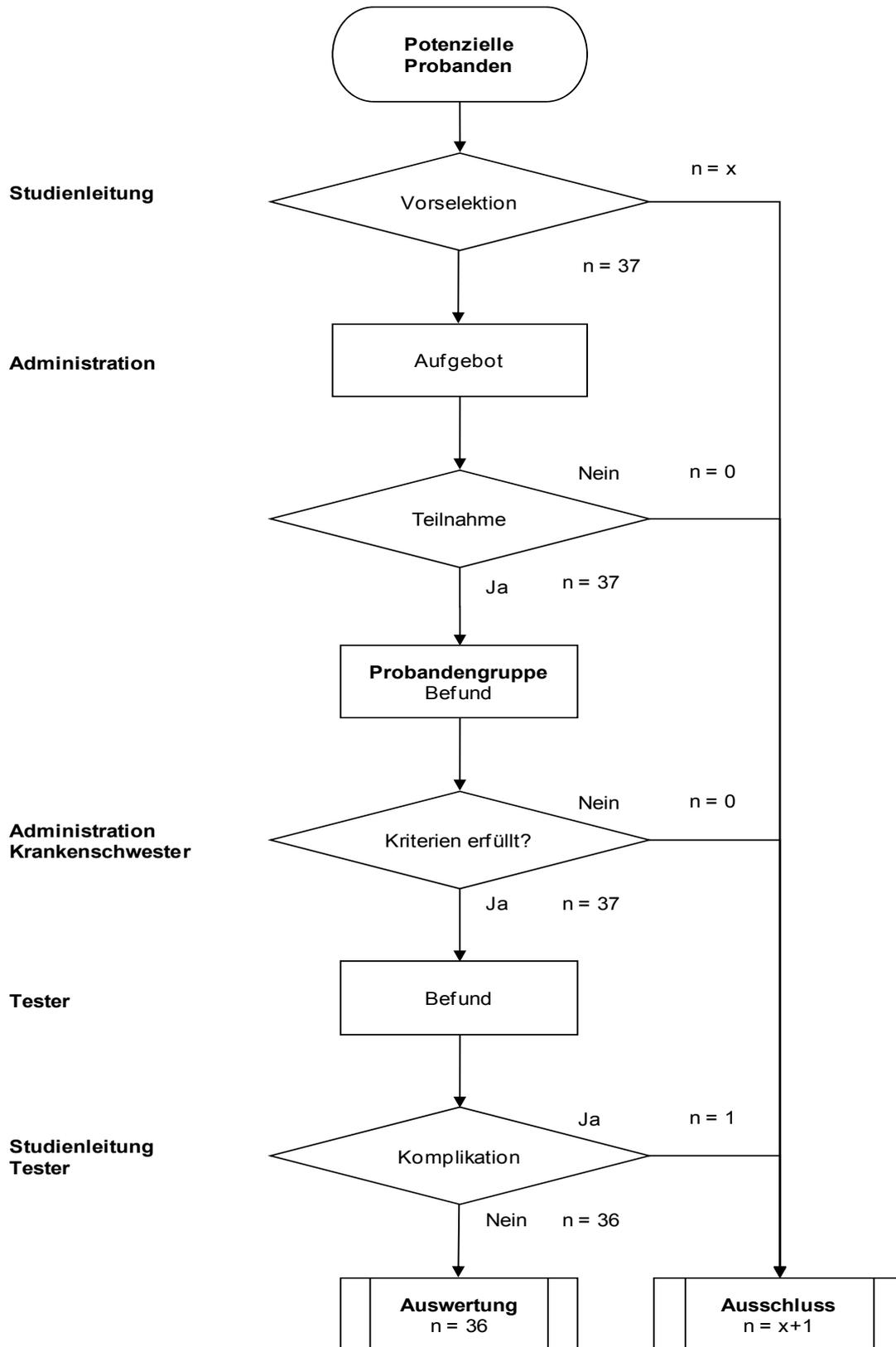


Abbildung 11 Flowchart Ein- und Ausschluss

Die Rekrutierung erfolgte aus dem Patientenpool des Osteoteams Burgdorf. Gefragt waren mindestens 30 Testpersonen. Um eventuelle Ausfälle abzudecken, wurden 37

Probanden rekrutiert. Nach Ermitteln der Ein- und Ausschlusskriterien konnten alle Probanden in die Studie eingeschlossen werden. Nach Abschluss der Studie fehlten bei einer Testperson die Daten eines Befundes. Dieser Proband wurde nachträglich von der Studie ausgeschlossen, da bei Anwendung der Holsti-Formel zur Berechnung des Koeffizienten keine Daten fehlen dürfen.

3.8.3. Stichprobe

Es konnte eine Probandengruppe mit folgender Zusammensetzung als Stichprobe ausgewertet werden:

Geschlecht	Weiblich	Männlich
Anzahl Probanden	23	13
Durchschnittsalter	39.99	45.12
Maximales Alter	80.02	74.64
Minimales Alter	15.22	12.57

Tabelle 3 Verteilung der Stichprobe nach Alter und Geschlecht

Die Altersstruktur der Stichprobe ist (wie aus Tabelle 3 ersichtlich) gut verteilt und erfüllt somit die Anforderungen der Homogenität.

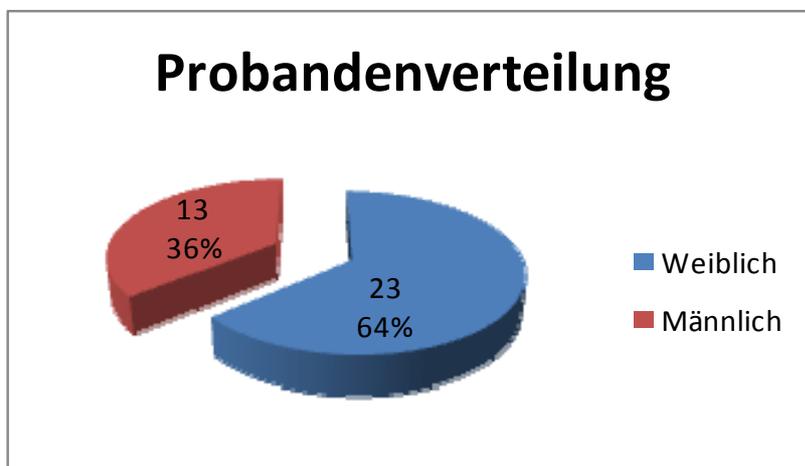


Abbildung 12 Anzahl und prozentuale Verteilung des Geschlechts

Der Anteil an weiblichen Probanden ist (wie aus Abbildung 12 ersichtlich) etwa doppelt so gross wie der Anteil der männlichen Probanden.

Der gesundheitliche Zustand der Probanden in Bezug auf deren Rückenschmerzen war wie folgt:

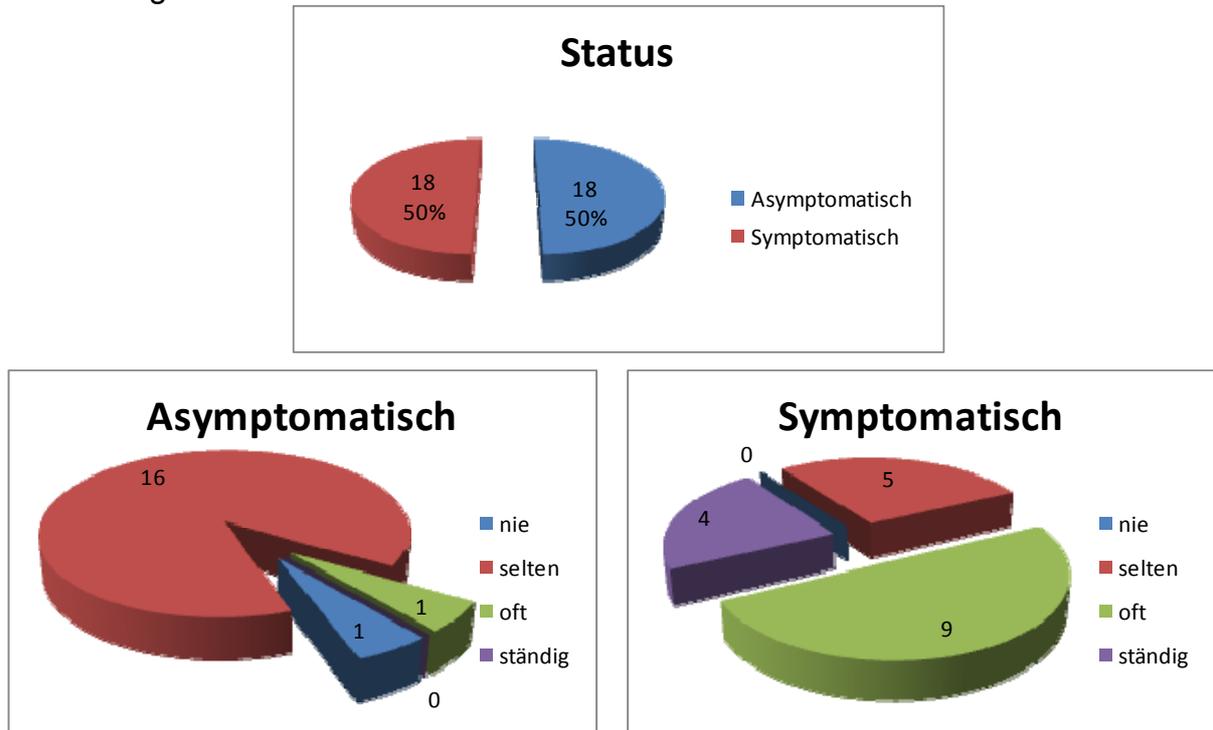


Abbildung 13 Verteilung der Schmerz-Symptomatik in der Stichprobe

In Bezug auf die Schmerz-Symptomatik liegt (wie in Abbildung 13 ersichtlich) eine gute Verteilung vor. Der Anteil Probanden mit Schmerzen ist gleich hoch wie derjenige ohne Schmerzen.

Die meisten Probanden (16 von 18) der asymptomatischen Gruppe (hatten bei der Testung keine Symptome) gaben an, selten Schmerzen zu haben.

Die Probanden der symptomatischen Gruppen (hatten bei der Testung Symptome) äusserten sich in Bezug auf die Häufigkeit ihrer Schmerzen wie folgt: ein Viertel gab an ständig an Schmerzen zu leiden, die Hälfte oft und ein Viertel selten.

Somit liegt bezüglich der relevanten Kriterien eine genügend homogene Stichprobe vor. Diese lässt mindestens auf die Population der Patienten in meiner Praxis anwendbare Schlüsse zu.

3.8.4. Durchführung der Analyse, Ablauf

Jeder Proband wurde vorgängig mündlich und schriftlich (siehe Anhang) durch die Studienleitung über die Studie und den Ablauf in Kenntnis gesetzt. Jeder Proband unterschrieb eine Einverständniserklärung zur Teilnahme an der Studie. Bei

Minderjährigen wurde das schriftliche Einverständnis der Eltern eingeholt. Die Probanden wurden entsprechend dem Zeitplan in Vierer-Gruppen eingeteilt und als Gruppe zur gleichen Zeit aufgeboten.

Von jedem Probanden werden gemäss Anamneseprotokoll (siehe Anhang) seine Daten durch die Sekretärin des Osteoteams (= Krankenschwester) erhoben und im PC festgehalten. Danach werden durch die Studienleitung bei allen Probanden die nötigen Sicherheitstests durchgeführt und die Dornfortsätze der Wirbelkörper mit kleinen selbstklebenden Etiketten markiert. In vier getrennten Räumen werden diese vier Probanden von den vier Osteopathen je ein Mal nach dem vorgegebenen Protokoll getestet. Die Befunde werden von Assistenten (kein medizinisch geschultes Personal) direkt in den PC eingegeben. Die Osteopathen haben keine Einsicht in die Befunde der Kollegen und in die Anamnese der Probanden. Die Koordination und Geheimhaltung der ermittelten Daten wird durch den Statistiker gewährleistet.

Vier Osteopathen, die seit mindestens 3 Jahren täglich nach der Methode der *LMO* ihren Befund erheben, testen je 36 Probanden zwei Mal. Für die Befunderhebung hat der Osteopath maximal fünf Minuten zur Verfügung. Der Befund erfolgt wie in 2.3.2 beschrieben. Bei jedem Probanden wird zuerst die *Partielle Totale Läsion* ermittelt und anschliessend die *Dominante Läsion* der Wirbelsäule bestimmt. Danach wechseln die Osteopathen im Uhrzeigersinn die Räume und untersuchen den nächsten Probanden. Der Rundgang wird zwei Mal wiederholt. Das bedeutet, dass jeder Osteopath in 40 Min. acht Befunde erheben muss und jeder Patient acht Mal getestet wird.

Alle 45 Min. wechselt die Probandengruppe. Es wurden insgesamt 9 Gruppen mit 37 Probanden geplant. Ein Proband fiel durch das Fehlen von Daten in der Erfassung aus.

Die Studie konnte wie geplant durchgeführt werden. Die Zeitplanung klappte einwandfrei. Alle Probanden erschienen pünktlich und motiviert. Es war ein schöner Tag, der dank dem Einsatz eines 16-köpfigen Hilfs-Teams und der Probanden erfolgreich über die Bühne ging.

3.8.5. Reaktion der Probanden auf den Untersuch

Zur Überprüfung vegetativer Reaktionen wurden die Probanden nach der Testung nach ihrem Befinden befragt. Die Probanden gaben verschiedene Symptome an. Bei allen gab es eine Hyperämie der Haut im Bereich der Testung, die rasch verschwand. Dies wurde vor der Abnahme der Klebeetiketten durch die Studienleitung überprüft. Gemäss Aussagen der Patienten und der Assistenten führten die Tester die Tests entgegen der vorherigen Absprache mehrmals durch (3.7). Zudem zeigte sich aus den Angaben der Patienten, dass die Tester die Tests mit sehr unterschiedlicher Druckstärke ausgeführt hatten. Folgende Reaktionen traten auf:

- 2 reagierten mit Schwitzen
- 1 reagierte mit Übelkeit
- 8 fühlten sich beweglicher
- 5 gaben mehr Schmerzen an
- 21 hatten keine Reaktion

Die Fragen inwieweit diese Befunderhebung, acht Testungen in 40 Minuten, einen Einfluss auf den sympathischen Grenzstrang hat und ob in den Facettengelenken eine Mobilisation stattfindet, scheinen berechtigt zu sein (siehe 3.7).

4. Ergebnisse

Der Befund der Wirbelsäule nach der *LMO-Methode* beinhaltet zwei Aspekte: das Ermitteln der *Partiellen Totalen Läsion* und das Bestimmen der *Dominanten Läsion*. Die Hauptergebnisse der Untersuchung lauten wie folgt:

Die Berechnungen des Reliabilitätskoeffizienten nach Holsti **bestätigten** die Hypothesen 1.1. und 1.2. für die Ermittlung der *Partiellen Totalen Läsion*.

- Hypothese 1.1.:
Bestätigt wird die Reproduzierbarkeit bezüglich der *Partiellen Totalen Läsion* für die Intra-Rater-Reliabilität mit einem Mittelwert aller Testungen von r_H : 0.794(\pm 0.098)
- Hypothese 1.2.:
Bestätigt wird die Reproduzierbarkeit bezüglich der *Partiellen Totalen Läsion* für die Inter-Rater-Reliabilität mit einem Mittelwert aller Testungen von r_H : 0.677(\pm 0.068).

Nicht bestätigt werden konnte nach Holsti die Übereinstimmung bei der Bestimmung der *Dominanten Läsion*.

Die Berechnungen des Reliabilitätskoeffizienten nach Holsti **bestätigten nicht** die Hypothesen 2.1. und 2.2. für die Bestimmung der *Dominanten Läsion*.

- Hypothese 2.1.:
Nicht bestätigt wird die Reproduzierbarkeit bezüglich der *Dominanten Läsion* für die Intra-Rater-Reliabilität mit einem Mittelwert aller Testungen von r_H : 0.465(\pm 0.501)
- Hypothese 2.2.:
Nicht bestätigt wird die Reproduzierbarkeit bezüglich der *Dominanten Läsion* für die Inter-Rater-Reliabilität mit einem Mittelwert aller Testungen von r_H : 0.167(\pm 0.193).

4.1. Ergebnisdarstellung

Die Ergebnisse, welche alle mit dem Programm PRAM berechnet wurden, sind für die vier Hypothesen in Diagrammen dargestellt. Sämtliche Diagramme wurden sowohl für die *Partielle Totale Läsion* als auch für die *Dominante Läsion* wie folgt erstellt.

4.1.1. Ergebnisse Intra-Rater

Bei den Intra-Rater-Ergebnissen, werden die Ergebnisse der beiden Durchgänge für jeden Tester einzeln ausgewertet und als arithmetisches Mittel mit der Standardabweichung im Diagramm dargestellt. Zusätzlich wird das arithmetische Mittel der vier Tester unter „Avg“ (Average) dargestellt.

4.1.2. Ergebnisse Inter-Rater

Bei den Inter-Rater-Ergebnissen, werden die Ergebnisse eines Durchganges für jedes mögliche Tester-Paar (sechs Paare) ausgewertet und als arithmetisches Mittel mit der Standardabweichung im Diagramm dargestellt. Dabei werden die Ergebnisse beider Durchgänge pro Tester-Paar dargestellt. Zusätzlich wird das arithmetische Mittel der sechs Tester-Paare für beide Durchgänge einzeln unter „Avg“ (Average) dargestellt.

4.1.3. Ergebnisse abhängig von Geschlecht und Symptom-Status

Für die Untersuchung des Einflusses von Geschlecht und Symptom-Status wurden die Ergebnisse ebenfalls in je drei analogen Diagrammen dargestellt:

- Intra-Rater-Reliabilität
- Inter-Rater-Reliabilität Durchgang 1
- Inter-Rater-Reliabilität Durchgang 2

4.2. Bestimmung der *Partiellen Totalen Läsion*

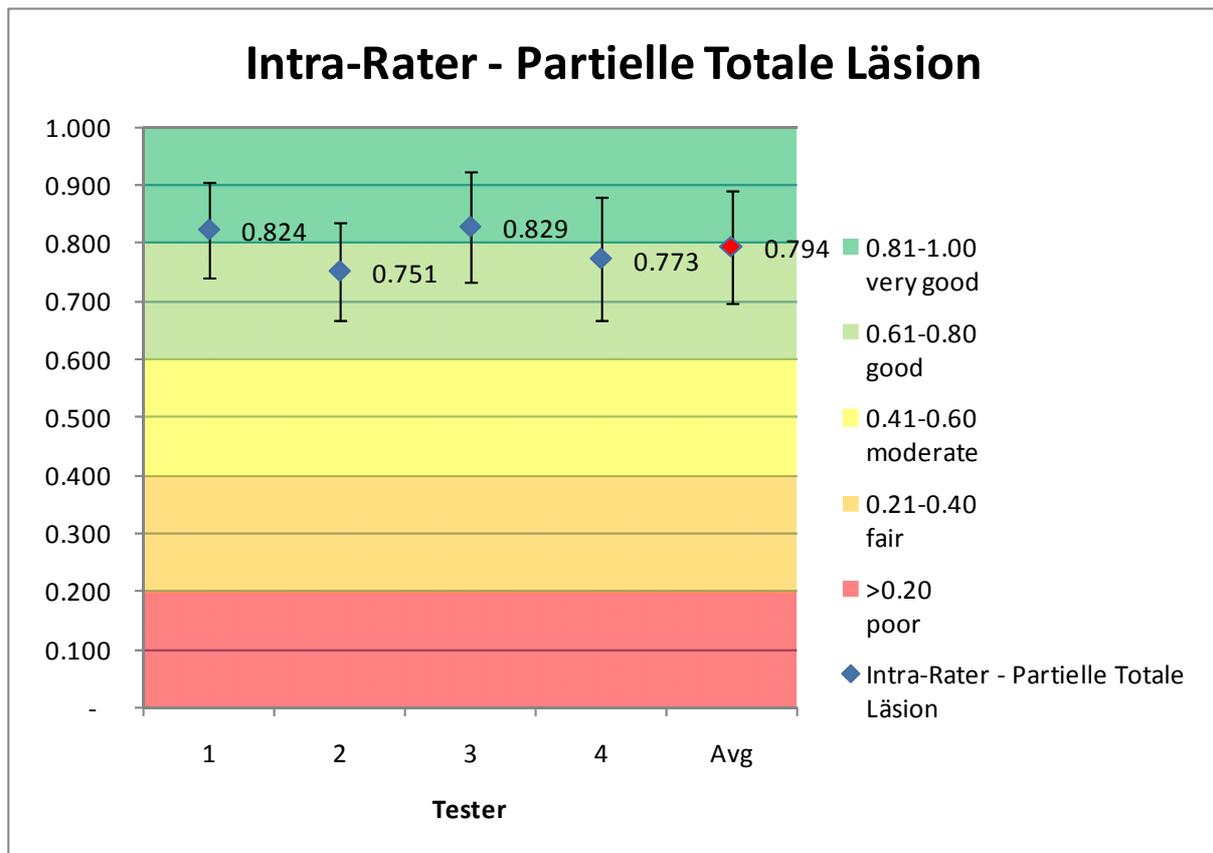


Abbildung 14 ICC Holsti's Intra-Rater – Partielle Totale Läsion

Avg = Arithmetisches Mittel der Ergebnisse der vier Tester

ICC = Reliabilitäts-Koeffizient nach Holsti (arithmetisches Mittel der Einzelergebnisse beider Durchgänge eines Testers)

Die Ergebnisse der Intra-Rater-Reliabilität für die Ermittlung der *Partiellen Totalen Läsion* befinden sich bei allen vier Testern im Bereich von gut bis sehr gut. Bei zwei Testern ist deren Mittelwert im Bereich über $r_H:0.8$, bei zwei Testern knapp darunter. Die mittlere Intra-Rater-Reliabilität liegt mit $0.79(\pm 0.098)$ knapp nicht im Bereich sehr gut. Die kleine Standardabweichung zeigt, dass 2/3 aller Werte der Grundgesamtheit wahrscheinlich im Bereich gut bis sehr gut liegen (Hüsler/Zimmermann, 2006: 58).

Die Hypothese 1.1. ist als bestätigt zu erachten. Es gibt eine gute Intra-Rater-Reliabilität für die Ermittlung der *Partiellen Totalen Läsion* der Wirbelsäule nach der *LMO-Methode*.

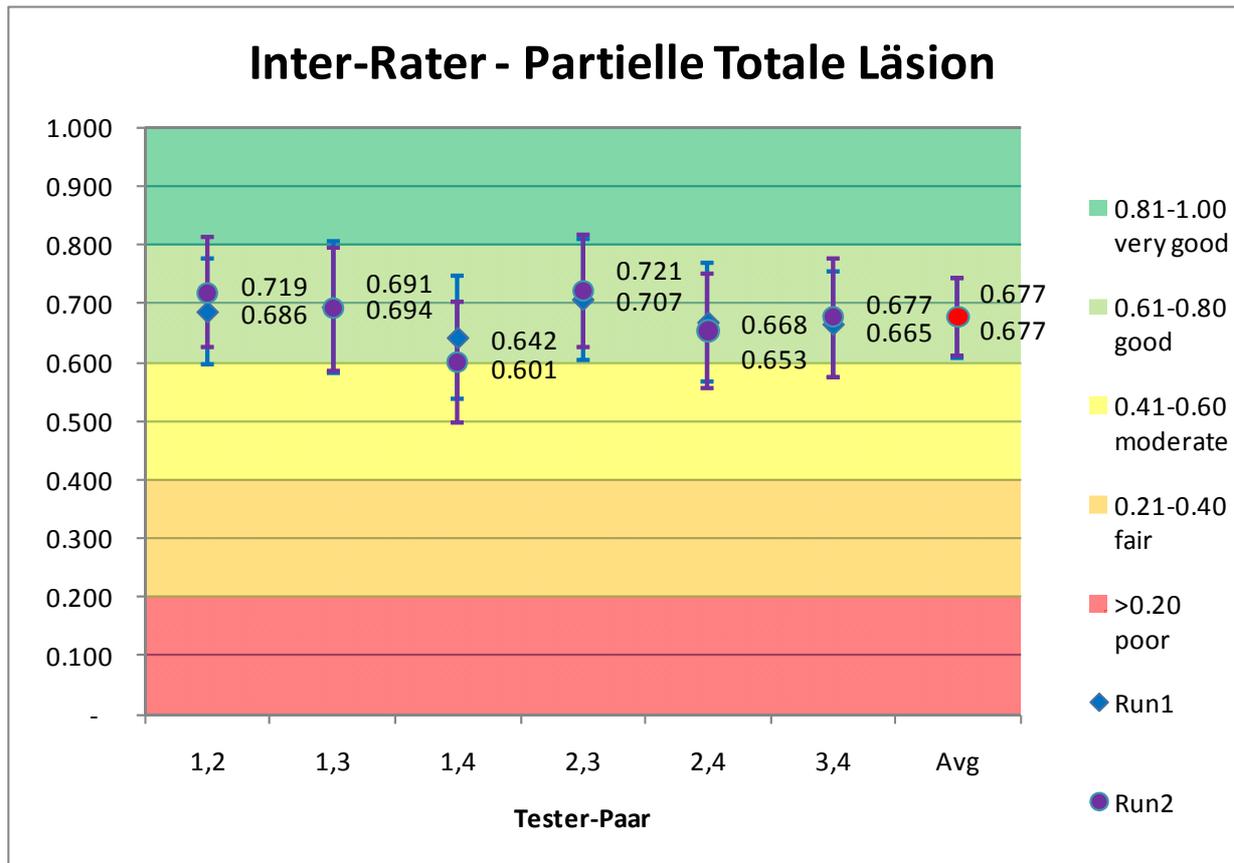


Abbildung 15 ICC Holsti's Inter-Rater – Partielle Totale Läsion

Avg = Arithmetisches Mittel der Ergebnisse aller sechs Tester-Paare (je Durchgang)

ICC = Reliabilitäts-Koeffizient nach Holsti (arithmetisches Mittel der Einzelergebnisse zweier Tester im selben Durchgang. Pro Tester-Paar zwei Ergebnisse)

Die Ergebnisse der Inter-Rater-Reliabilität für die Ermittlung der *Partiellen Totalen Läsion* befinden sich bei allen vier Testern im Bereich gut. Die mittlere Inter-Rater-Reliabilität beträgt $0.68(\pm 0.068)$. Die kleine Standardabweichung zeigt, dass 2/3 aller Werte der Grundgesamtheit wahrscheinlich im Bereich gut liegen (Hüsler/Zimmermann, 2006).

Die Hypothese 1.2. ist als bestätigt zu erachten. Es gibt eine gute Inter-Rater-Reliabilität für die Ermittlung der *Partiellen Totalen Läsion* der Wirbelsäule nach der *LMO-Methode*.

4.3. Normalverteilung der Ergebnisse

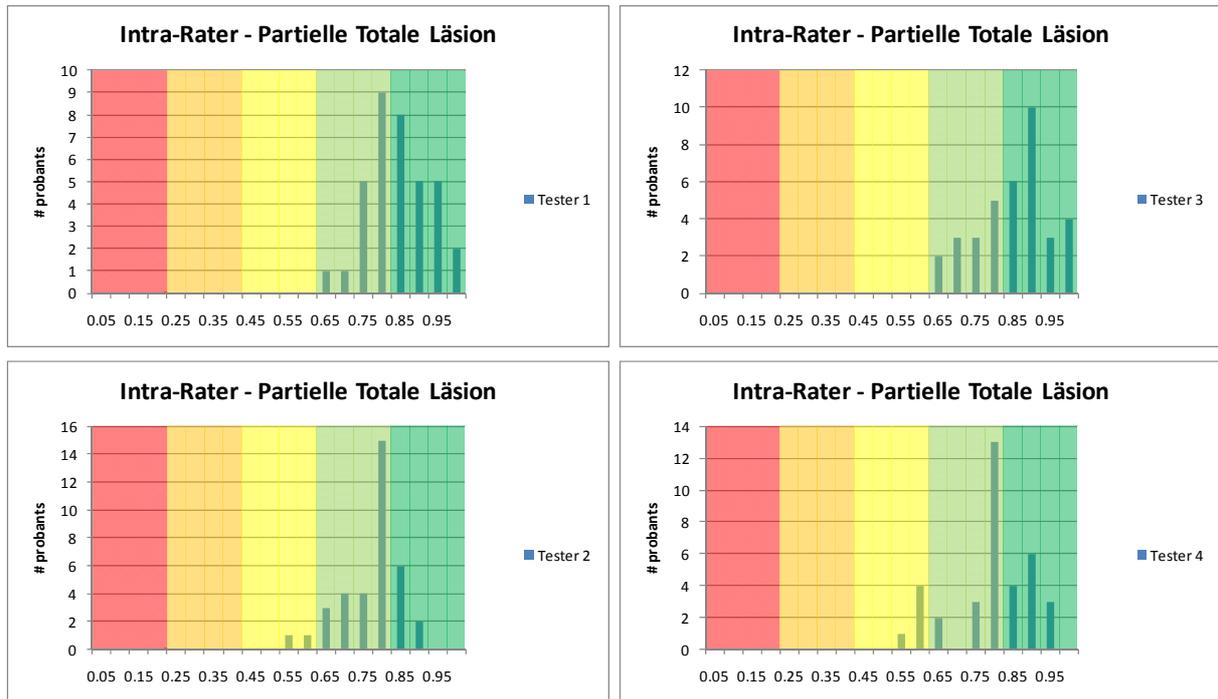


Abbildung 16 ICC Holsti's Intra-Rater – Partielle Totale Läsion – Verteilung der Ergebnisse

Die optische Beurteilung der Histogramme zeigt eine annähernd normalverteilte Stichprobe. Auf Empfehlung des Statistikers wurde auf eine weitergehende Untersuchung (z.B. statistische Tests der Schiefe) verzichtet, da die Bandbreite der Mittelwerte ($AVG \pm SDEV$) bei diesen Ergebnissen nicht mehr als zwei Klassen umfasst, also 2/3 aller Werte höchstens zwei Klassen umfassen. Die Genauigkeit der Berechnung liegt damit in jedem Fall höher als die Unsicherheit der Klassierung der Ergebnisse an sich (siehe 3.5.2).

4.4. Bestimmung der *Dominanten Läsion*

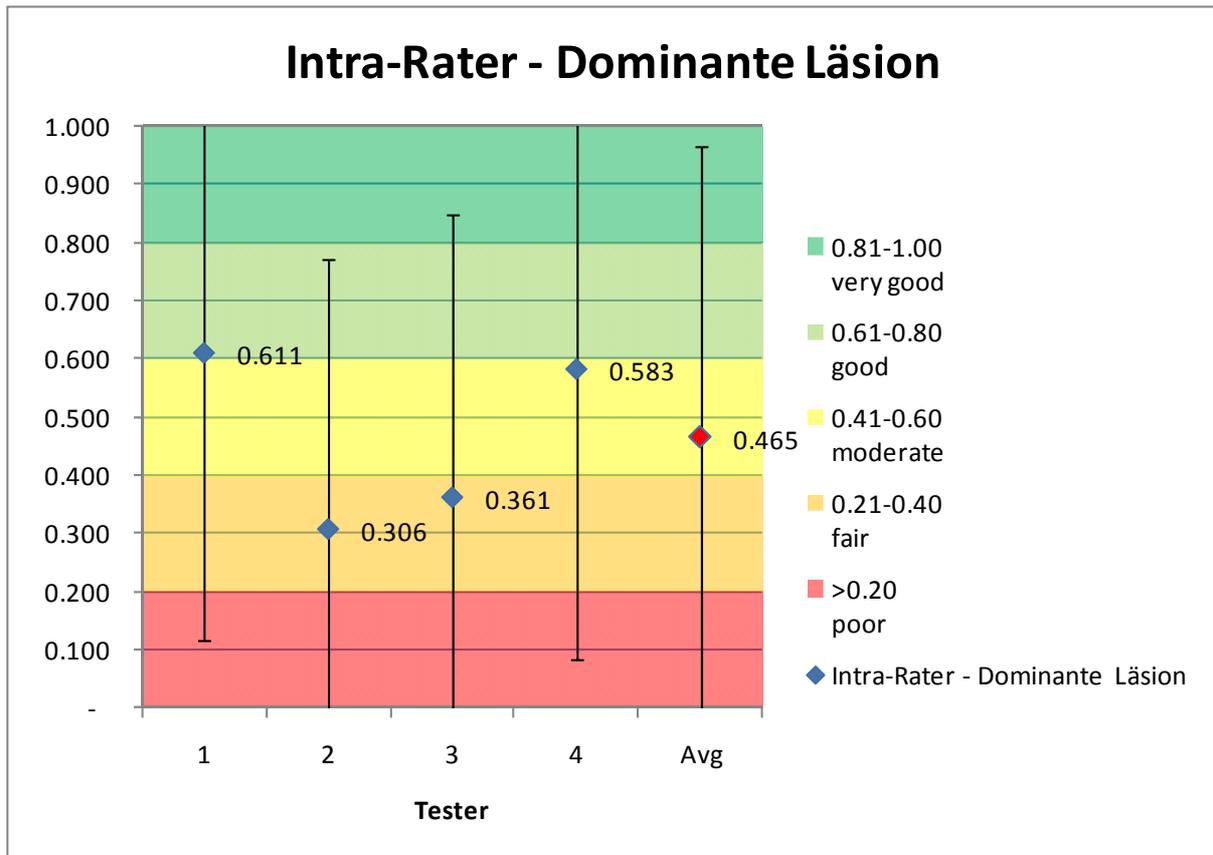


Abbildung 17 ICC Holsti's Intra-Rater – *Dominante Läsion*

Avg = Arithmetisches Mittel der Ergebnisse der vier Tester

ICC = Reliabilitäts-Koeffizient nach Holsti (arithmetisches Mittel der Einzelergebnisse beider Durchgänge eines Testers)

Die durchschnittlichen Ergebnisse der Intra-Rater-Reliabilität für die Bestimmung der *Dominanten Läsion* aller vier Tester befinden sich im Bereich moderat. Bei zwei Testern ist deren Mittelwert im Bereich um r_H : 0.6 bei zwei Testern um r_H : 0.3.

Die mittlere Intra-Rater-Reliabilität liegt mit r_H : 0.47 (± 0.501) ebenfalls im Bereich moderat. Die grosse Standardabweichung zeigt aber, dass 2/3 aller Werte der Grundgesamtheit wahrscheinlich über alle Klassen verteilt liegen und somit nicht aussagekräftig sind (Hüsler/Zimmermann, 2006: 58).

Die Hypothese 2.1. ist als nicht bestätigt zu erachten. Es gibt eine moderate Intra-Rater-Reliabilität für die Ermittlung der *Dominanten Läsion* der Wirbelsäule nach der *LMO-Methode*. Diese ist aber nicht aussagekräftig, da die Streuung zu gross ist.

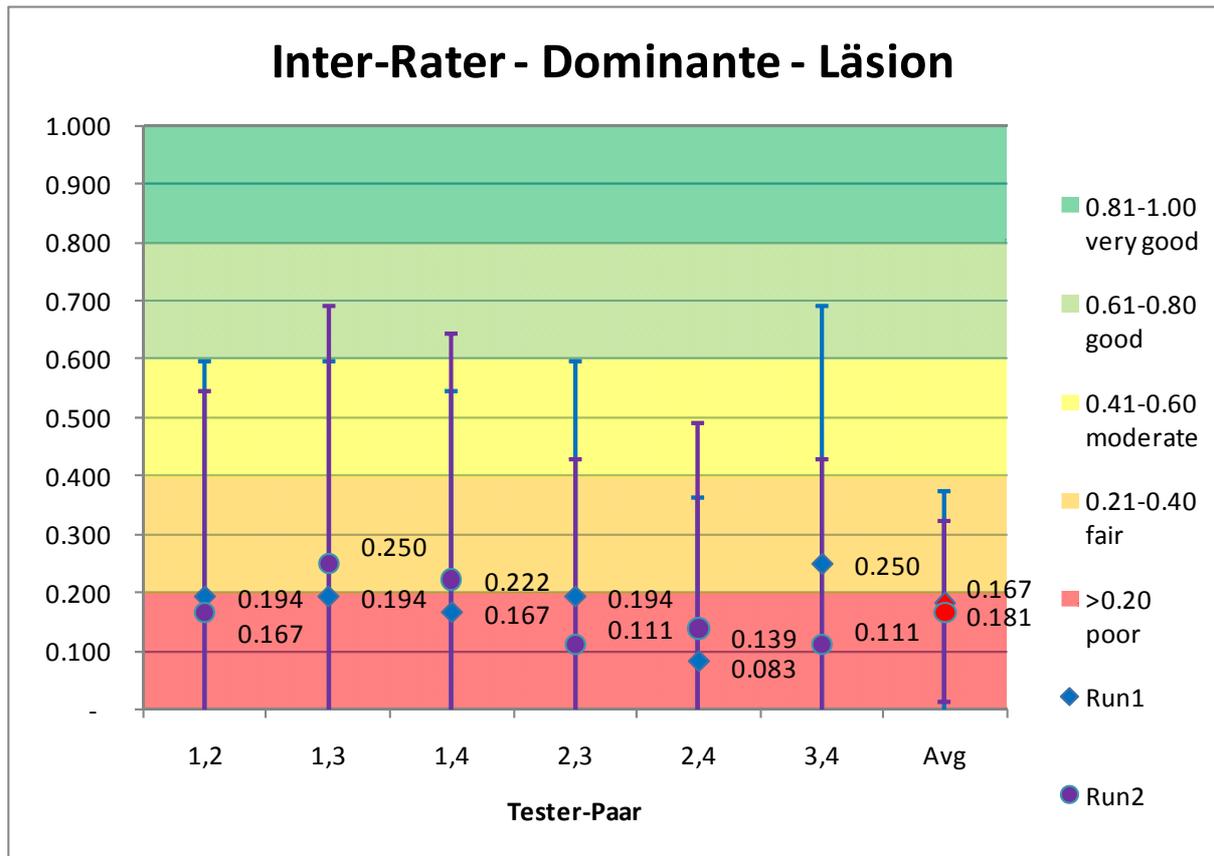


Abbildung 18 ICC Holsti's Inter-Rater – Dominante Läsion

Avg = Arithmetisches Mittel der Ergebnisse aller sechs Tester-Paare (je Durchgang)

ICC = Reliabilitäts-Koeffizient nach Holsti (arithmetisches Mittel der Einzelergebnisse zweier Tester im selben Durchgang. Pro Tester-Paar zwei Ergebnisse)

Die durchschnittlichen Ergebnisse der Inter-Rater-Reliabilität für die Bestimmung der *Dominanten Läsion* aller vier Tester befinden sich im Bereich nicht brauchbar.

Die mittlere Inter-Rater-Reliabilität liegt mit r_H : 0.18 (± 0.191) ebenfalls im unbrauchbaren Bereich. Die grosse Standardabweichung zeigt, dass 2/3 aller Werte der Grundgesamtheit wahrscheinlich in den zwei tiefsten Klassen liegen und somit nicht aussagekräftig sind (Hüsler/Zimmermann, 2006).

Die Hypothese 2.2. ist als nicht bestätigt zu erachten. Es gibt keine Inter-Rater-Reliabilität für die Ermittlung der *Dominanten Läsion* der Wirbelsäule nach der *LMO-Methode*.

4.5. Einfluss Status – Symptomatisch – Asymptomatisch

Es konnten keine signifikanten Unterschiede der Resultate zwischen erstem und zweitem Durchgang in Abhängigkeit des Status festgestellt werden.

Die ermittelten Koeffizienten unterscheiden sich unwesentlich zwischen Intra- und Inter-Rater (siehe Einfluss Status – Symptomatisch – Asymptomatisch 8.3).

4.6. Einfluss Geschlecht

Auch der Einfluss des Geschlechts auf das Resultat zeigte keine wesentlichen Unterschiede.

Die ermittelten Koeffizienten unterscheiden sich unwesentlich zwischen Intra- und Inter-Rater (siehe 8.4).

5. Diskussion

Überprüft wurde die Intra-und Inter-Rater-Reliabilität des Befundes der Wirbelsäule nach der *LMO-Methode*. Aus der *ersten Funktionellen Einheit* wurden mit dem *Provokationstest* die Wirbelkörper getestet. Der Befund beinhaltet folgende zwei Teile, das Ermitteln der *Partiellen Totale Läsion* und das Bestimmen der *Dominante Läsion* (siehe 2.2.). Die Hypothese der Reproduzierbarkeit der **Partiellen Totalen Läsion** mit dem *Provokationstest* konnte in beiden Fällen **bestätigt werden (siehe 4.2)**. Die Reproduzierbarkeit der **Dominanten Läsion** mit dem *Balancetest* konnte in beiden Fällen **nicht bestätigt werden (siehe 4.4)**. Dabei ist wichtig zu wissen, dass diese Einstufung keinen Anspruch auf Absolutheit hat. Die entsprechenden Resultate können jedoch im Vergleich zu Ergebnissen vergleichbarer Studien (manuelle Befunderhebungen der Wirbelsäule) durchaus eine Aussagekraft haben.

5.1. Reproduzierbarkeit der Partiellen Totalen Läsion

Die *Partielle Totale Läsion* gibt nach der *LMO-Methode* die Summe aller gefundenen *Dysfunktionen* im getesteten Bereich an. Das Ziel jeder Behandlung nach der *LMO-Methode* ist das Finden und Behandeln der *Partielle Totale / Totale Läsion*. Dies ist das Kernstück der Methode. Bei der Berechnung der Koeffizienten wurden die Übereinstimmungen der positiv und negativ kodierten *Provokationstests* berücksichtigt. Denn auch das Erkennen des Nichtvorhandenseins einer *Dysfunktion* gilt als Übereinstimmung.

5.1.1. Reproduzierbarkeit der Intra-Rater-Reliabilität

Die Reproduzierbarkeit des Wirbelsäulenbefundes scheint bei dieser Studie für die Intra-Rater-Reliabilität mit einem durchschnittlichen Wert um $r_H:0.8$ gut gegeben zu sein. Über die Bewertung solcher Koeffizienten streitet sich die Fachwelt (siehe 3.5.2) Aber für alle Spezialisten gilt ein Wert ab 0.8 als gut. (Auch wenn wie bei der Berechnungs-Methode nach Holsti der Faktor der zufälligen Übereinstimmungen nicht berücksichtigt ist.) Der bei der Studie erreichte Wert hört sich folglich vielversprechend an und sollte weiter untersucht werden, zumal solche Werte bei manuellen Methoden selten zu sehen sind (siehe 5.3).

Die Anzahl der Probanden ist ausreichend, um eine Aussage machen zu können. Sie ist aber an der unteren Grenze. Da die Ergebnisse gut waren, würde es sich

lohen, die Studie im grösseren Rahmen mit 200-300 Probanden (siehe 3.5.3) durchzuführen. Die Probanden waren zum grössten Teil eine Auswahl aus dem Patientenpool des Osteoteams. Entsprechend den relevanten Kriterien war diese Gruppe ausreichend homogen durchmisch (siehe 3.8.3). Dies lässt also mindestens Schlüsse auf die Population der Patienten der Praxis osteoteam.ch zu. Um allgemeingültige Schlüsse aus den Ergebnissen ziehen zu können, ist die Auswahl der Stichprobe nicht ausreichend. Hierzu müsste die Ermittlung der Stichprobe entsprechend aufwändiger erfolgen und nach den diesbezüglichen Kriterien (Hüsler/Zimmermann, 2006) durchgeführt werden.

Die Methode erhebt den Anspruch, genau und strukturiert zu sein. Die Frage, die zur Hypothese geführt hat, war, ob die Standardisierung für eine Intra-Rater-Reproduzierbarkeit des Wirbelsäulenbefundes ausreiche. Die Ergebnisse machen den Anschein, als wäre dies für den Befund der Wirbelsäule der Fall. Um eine Aussage in Bezug auf die Gültigkeit dieser Ergebnisse machen zu können, müssten diese in weiteren Studien bestätigt werden.

5.1.2. Reproduzierbarkeit der *Inter-Rater-Reliabilität*

Die Reproduzierbarkeit des Wirbelsäulenbefundes bei mehreren Testern ist mit einem Wert um $r_H:0.7$ noch akzeptabel. Es lässt sich, wie schon im statistischen Teil (siehe 3.5.2) beschrieben, darüber streiten, ab wann ein Wert aussagekräftig ist. Aber 0.7 ist eindeutig mehr als der Zufallswert 0.5. und das Fehlen der Korrektur des Zufallfaktors bei der Holsti-Formel kann wegen der vier Tester relativiert werden (siehe 3.5.3) Aus diesem Wert ist lesbar, dass es mindestens eine tendenzielle Reproduzierbarkeit gibt.

Interessant ist hier die Frage, weshalb ein Unterschied zwischen der Intra-Rater- und der Inter-Rater-Reliabilität vorliegt. Folgende Ursachen könnten einen Einfluss gehabt haben:

- Nicht erkannte Fehlerquellen (Tester wurden ungenügend geschult),
- Tester hielten sich nicht vollständig an die Vorgaben,
- Die Standardisierung der Methode ist bei einer Inter-Rater-Reliabilität nicht ausreichend oder nur knapp ausreichend, um das subjektive Empfinden des Testers bei der Interpretation der *Elastizität* wett zu machen.

5.1.3. Reproduzierbarkeit des Wirbelsäulenbefundes

Die Ergebnisse haben mich in Staunen versetzt. Nach dem Literaturstudium über Reliabilitätsstudien zu Befundmethoden der Wirbelsäule (5.3), bei dem ich keine brauchbaren Ergebnisse für manuelle Befundmethoden seit 1965 gefunden habe, rechnete ich nicht mit diesen Daten. Dies veranlasste mich, die Daten und Berechnungen mehrmals mit dem Statistiker zu überprüfen um auszuschliessen, dass sich Fehler eingeschlichen haben. Da dies nach unserem Ermessen nicht der Fall ist, suchte ich nach Faktoren, die eventuell zu diesem Ergebnis geführt haben.

5.1.4. Handhabung und Standardisierung der Testung

Bei den meisten Studien ist einer der diskutierten Fehlerfaktoren die fehlende Standardisierung des Befundablaufes und die Komplexität in der Beurteilung der Testung. In der Masterthesis von Dornieden (2008) werden genau diese Problem erwähnt. Sowohl die Inter-Rater- als auch die Intra-Raterreliabilität ergaben keine brauchbare Reproduzierbarkeit, selbst bei Testern, welche die Methode seit längerer Zeit anwenden und geübt sind. Im Gegenteil zeigten die Ergebnisse, dass die Intra-Raterreliabilität bei Testern frisch aus der Schule eher besser war. Als mögliche Fehlerquellen wurden die unterschiedliche Handhabung des Tests und mangelnde Standardisierung angegeben. Ein wichtiger Faktor bei der Reliabilitätsprüfung einer Messmethode ist, dass die Messung möglichst immer gleich erfolgt (siehe 2.4.4). In unserem Fall wäre dies die Ausführung des Tests. Die möglichst uniforme und präzise Handhabung eines Tests ist ein wichtiger Faktor für die Reproduzierbarkeit einer Methode. Die Tester in dieser Studie arbeiten täglich mit der *LMO-Methode* und sind darin geübt. Bei jedem Patienten unabhängig von seinen Beschwerden wird die Testung der Wirbelsäule durchgeführt. Die Tester sind auch Mitglieder einer *LMO-Übungsgruppe*, welche sich seit zwei Jahren im Abstand von zwei Monaten trifft. Bei diesen Treffen werden unter anderem das Handling und der Ablauf des Befundes geübt und aufgefrischt. Zusätzlich gehören zwei der Mitglieder dieser Gruppe dem Lehrerteam von Chauffour und Pratt an und sind deshalb oft mit dem Vermitteln der präzisen Ausführung der Tests nach der *LMO-Methode* beschäftigt. Es ist davon auszugehen, dass die Tester einen ungewöhnlich hohen Grad an übereinstimmendem Handling vorweisen können. Dies könnte ein wesentlicher Faktor für die gute Reliabilität in dieser Studie sein.

Der *LMO-Befund* erhebt den Anspruch, hoch standardisiert zu sein. Dabei sind meiner Meinung nach zwei Faktoren von Bedeutung. Erstens ist der Test sehr einfach in der Anwendung und Durchführung. Er erfordert keine hochkomplexen dreidimensionalen Einstellungen, um durchgeführt werden zu können und ist entweder positiv oder negativ. Diese Einfachheit lässt wenig Spielraum für Eigenkreativität offen. Je einfacher ein Test ist, desto einfacher ist das Erlernen des gleichen Handlings durch verschiedene Tester. Zweitens erfolgt der Ablauf der Testung immer nach demselben Protokoll und wiederholt sich bei jeder Befunderhebung. Dies könnte ein weiterer Faktor für die guten Ergebnisse dieser Studie sein.

Bei dieser Studie wurden nicht nur die positiven Tests als übereinstimmend gewertet, sondern auch die negativen, weil diese bewusst negativ codiert wurden. Das ergab insgesamt 25 Tests pro Patient und Durchgang. Beim Überprüfen, ob die Ergebnisse richtig ermittelt wurden, weil sie uns zu gut vorkamen, stellten wir fest, dass sie ganz anders ausgesehen hätten, wenn wir nur die positiven Tests in die Berechnung eingeschlossen hätten. Bei den meisten Studien wurden nur die positiven gewerteten Tests erfasst. Vielleicht sollten auch die negativen in die Berechnungen einbezogen werden, vorausgesetzt dass sie bewusst als negativ getestet wurden.

5.1.5. Einfluss der Sensibilität des Testers auf das Ergebnis

Bei der Intra-Rater-Reliabilität wird der Befund vom selben Tester beim selben Probanden mehrmals durchgeführt. Bei dieser Art von Testung fällt der Faktor der subjektiven Empfindung des Testers nicht ins Gewicht, da immer derselbe Tester entscheidet. Im Unterschied dazu wird die Testung bei der Inter-Rater-Reliabilität von verschiedenen Testern mit verschiedener Sensibilität durchgeführt. Den Einfluss dieses Faktors versuchte ich im Vorfeld mit der Massnahme, dass jeder Osteopath seit mindestens drei Jahren nach dieser Methode arbeitet, einzudämmen. Ich ging davon aus, dass ein Unterschied in der Ausführung des Tests punkto Druck und Annäherung wegen der Erfahrung der Tester mit der Methode vernachlässigt werden kann. Deshalb verzichtete ich auf eine vertiefte Schulung der Tester zum Abstimmen des Drucks bei der Ausführung des Tests. Das könnte ein Fehler gewesen sein, wie die Befragung der Probanden nach der Testung ergab. Es stellte sich heraus, dass der Druck, der bei der Testung ausgeübt wurde, von Tester zu Tester stark variierte. Auffällig bei der Auswertung der Ergebnisse war auch, dass 2 Tester deutlich mehr

positive *Provokationstests* hatten als die beiden anderen. Dies könnte ein Grund für den Unterschied der Ergebnisse Inter- und Intra-Rater sein. Interessant ist auch, dass die zwei Tester, bei denen die Intra-Rater-Übereinstimmung höher als 0.8 war, bei der Inter-Rater-Übereinstimmung die grösste Differenz aufwiesen. Der eine dieser beiden Tester hatte die geringste Zahl an positiven Tests, der andere die grösste Anzahl an positiven Tests. Der Unterschied war vorwiegend auf die grössere Anzahl an positiven Tests zurückzuführen. Laut Aussage der Probanden hatte der Tester mit der grössten Anzahl positiver Kodierungen auch mit Abstand den stärksten Druck und umgekehrt. Daraus könnte man schliessen, dass der ausgeübte Druck einen Einfluss auf die Interpretation des Testes hat. Dies spricht dafür, sich Zeit zu nehmen, den Druck mittels Schulung der Tester anzugleichen. Mit dieser Massnahme würde der Druck-Faktor als unberechenbarer Faktor entschärft und die subjektive Empfindung könnte dadurch weniger ins Gewicht fallen. Es ist anscheinend nicht ausreichend, nur Erfahrung mit der Methode zu haben, um stabile Rahmenbedingungen zu gewährleisten. Die Fehleranalyse sollte um diesen Punkt ergänzt werden.

5.1.6. Reproduzierbarkeit der *Dominante Läsion*

Die Reproduzierbarkeit der *Dominanten Läsion* beim Befund der Wirbelsäule ergab weder bei der Intra- noch bei der Inter-Rater-Reliabilität brauchbare Ergebnisse. Laut Chauffour (2005) dient das Bestimmen der *Dominanten Läsion* der Abkürzung der Behandlung. Wie schon erwähnt geht das *LMO-Modell* davon aus, dass *Läsionen* einander dominieren können. Normalisiert man eine *Läsion*, so verschwinden die damit gekoppelten *Läsionen* auch. Ich wende dieses Verfahren in meiner Praxis an, und da scheint es zu funktionieren. Nach den Ergebnissen scheint der *Balancetest* eine Momentaufnahme zu sein. Das Ergebnis konnte selbst vom selben Tester nicht reproduziert werden. Es stellt sich die Frage, ob die Rahmenbedingungen für diesen Test zu unstetig waren (Einfluss der Anzahl Testungen) oder ob der *Balancetest* zu wenig standardisiert ist. Die *LMO-Methode* plädiert dafür, die *Primäre Läsion* zu finden. Ob das möglich ist, ist unter den gegebenen Umständen fragwürdig und bedarf weiterer Untersuchungen (siehe 5.1.5). Da das Bestimmen der *Dominanten/Primären Läsion* dazu dient, die Behandlung der *Totalen Läsion* möglichst effizient zu gestalten, ist deren Wichtigkeit für die eigentliche Befunderhebung von untergeordneter Bedeutung.

Laut Statistiker wäre für die Bestimmung der *Dominanten Läsion* (ein Messwert mit 25 möglichen Ausprägungen je Proband) eine vertiefte Analyse der Signifikanz z.B. mit einem T-Test (Hüsler/Zimmermann, 2006) sinnvoll gewesen. Aufgrund der geringen Übereinstimmung der Messwerte wurde jedoch darauf verzichtet.

5.1.7. Verändert die Testung der Facettengelenke die Rahmenbedingungen?

Beim Erstellen der Fehleranalyse stellte sich die Frage, inwieweit die acht Testungen punkto Mobilität der Wirbelsäule und vegetativer Reaktionen (siehe 3.7) die Rahmenbedingungen von Durchgang zu Durchgang verändern. Aus diesem Grunde wurde vorgegeben, je Durchgang und Wirbelkörper den *Provokationstest* nur einmal durchzuführen. Bei Zweifel, ob der Test positiv oder negativ ist, soll der Test nicht wiederholt sondern als negativ kodiert werden. Alle vier Tester hielten sich nicht an diese Abmachung. Bei Unsicherheit über die Kodierung wurde der Test mehrmals wiederholt. Der Grund dieses Handelns, war laut Tester der Wunsch, den Auftrag möglichst gut zu erfüllen. Für die Tester war die Testsituation wie eine Prüfungssituation. Entsprechend war der Druck, der auf ihnen lastete. Anscheinend habe ich den psychischen Stressfaktor, der durch das Testen verursacht wurde, unterschätzt. Diesen Faktor könnte man mit Trainingseinheiten relativieren. Die Studienleitung hätte vorab mehrere praktische Übungseinheiten durchführen sollen, bei denen der Ablauf der Testung hätte automatisiert werden können. Dieses Üben hätte den Testern mehr Sicherheit gegeben und so die Situation während der Testung entspannt. Mit diesen Voraussetzungen, das heisst mehr Testungen pro Wirbelkörper, wäre nun anzunehmen, dass sich die Rahmenbedingungen verändern würden.

Um eine Aussage betreffend die Rahmenbedingungen zu machen, verglich ich die durchschnittlichen Koeffizienten des ersten Durchganges mit denen des zweiten Durchganges. Falls die acht Testungen der Facettengelenke einer Behandlung dieses entsprächen, wäre zu erwarten, dass sich die Ergebnisse aus dem ersten und zweiten Durchgang unterscheiden. Zum Beispiel durch weniger positive Kodierungen, andere Kodierungen und so weiter. Nun war ja die Anzahl Testungen pro Wirbelkörper sogar erheblich höher als acht. Die Ergebnisse aus dem ersten und dem zweiten Durchgang sind aber praktisch identisch. Also kann davon ausgegangen werden, dass die Testung der Facettengelenke die

Rahmenbedingungen bezüglich deren Elastizität nicht verändert. Die Annahme, dass durch die Testung eine Behandlung erfolgt, ist bei diesem Ergebnis fragwürdig. Dies ist aber weiter zu untersuchen. Chauffour ging davon aus, dass der *Provokationstest* nicht ausreicht, um eine Veränderung der Elastizität in den Facettengelenken zu bewirken. Er könnte mit seiner Annahme recht haben. Um schlüssige Aussagen machen zu können, müsste dies aber näher untersucht werden.

Ob die Testung wegen der möglichen Mobilisation der Facettengelenke einen Einfluss auf die Rahmenbedingungen in Bezug auf die Ermittlung der *Dominanten Läsion* hatte, ist also offen. Um diesen Einfluss bestimmen zu können, müsste eine Intra-Rater-Reliabilitäts-Studie durchgeführt werden. Bei dieser Studie müsste die Anzahl der Testungen auf ein Minimum beschränkt werden (nur zwei Testungen pro Wirbelkörper).

Die Probanden wiesen vegetative Reaktionen auf (siehe 3.8.5). Bei der Fehleranalyse stellte sich die Frage, inwieweit die Ventralisierung der Wirbelkörper durch die Testung den Grenzstrang stimuliert. Ob die vegetativen Reaktionen wirklich durch die Stimulation des vegetativen Grenzstrangs mittels Ventralisation der Wirbelkörper verursacht wurde, kann nicht schlüssig beantwortet werden. Aber es steht fest, dass es vegetative Reaktionen gab. Auch hierzu müssten weitere Untersuchungen durchgeführt werden.

5.2. Wahl des Koeffizienten

Die Wahl des richtigen Koeffizienten für diese Art von Studie erwies sich bei der Vielzahl an Berechnungsvarianten als sehr schwierig (Rössler, 2005). Die Formeln, welche zur Berechnung des Reliabilitätskoeffizienten entwickelt wurden, wurden besonders für Inhaltsanalysen (Kommunikationswissenschaft) entwickelt und nicht für klinische Untersuchungen. Die klinische Forschung wendet jedoch bei Reliabilitätstestungen einer Methode dieselben Formeln an. Alle Berechnungsmodelle basieren auf dem Paarvergleich. Es gibt keine direkten Berechnungen, bei denen mehrere Tester berücksichtigt werden. Die Formeln werden bei nicht nominalskalierten Variablen angewandt und müssen entsprechend bei Nominalskalierung (z.B. Variable mit den Werten Ja/Nein) angepasst werden. Die Vor- und Nachteile der verschiedenen Formeln mussten deshalb gut abgewogen

werden. Von Rössler (2005: 190) wird für die Forschungspraxis bei kleinen Studien die Berechnung nach Holsti empfohlen. Die Argumente, die gegen diese Formel sprechen und berücksichtigt werden mussten, sind in 0 aufgelistet. Bei meiner Arbeit handelt es sich um eine Studie mit Nominalskalierung. Es gibt keine fehlenden Werte (falls das der Fall ist, scheidet der Proband aus). Die Variablenausprägung ist zwei. Somit müssen die Ergebnisse laut Rössler deutlich über 0.5 liegen, um über einer zufälligen Übereinstimmung zu liegen. Die fehlende Korrektur für die zufällige Übereinstimmung wird bei der Inter-Rater-Reliabilität relativiert, da vier Tester untersuchen und so der Zufallsfaktor wie in 3.5.2. beschrieben von 50% (0.5) auf 3.1% (0.031) verändert wird. So sind alle Kritikpunkte nach Hayes/Krippendorff entschärft, respektive die Kritikpunkte (ausser dem Zufallsfaktor bei der Intra-Rater-Reliabilität) entsprechen den Anforderungen der Studie. Nur bei der Intra-Rater-Reliabilität müssen wir die fehlende Korrektur für zufällige Übereinstimmungen in Kauf nehmen. Das bedeutet, dass bei der Inter-Rater-Reliabilität der Koeffizient 0.8 erreichen sollte. Mit dem durchschnittlichen Wert von 0.79 ist dies erreicht. Die Holsti-Formel scheint für diese Untersuchung die geeignetste zu sein, da sie genau den Anforderungen entspricht. Die Klassierung der Ergebnisse hat, wie erwähnt, keinen Anspruch auf Absolutheit.

Die Frage nach der geeigneten Berechnungsmethode ist nicht abschliessend geklärt. Um bei solchen klinischen Studien verlässliche, interpretierbare und vergleichbare Ergebnisse zu erhalten, müssten die verschiedenen Berechnungsmodelle vertieft evaluiert werden. Bislang gibt es kaum Reliabilitätsstudien für manuelle Wirbelsäulenuntersuchungen, die den evidenzbasierten Qualitätsansprüchen genügen (siehe 5.3). Vielleicht ist die Wahl des Berechnungsmodells mit ein Grund für diesen Mangel, da das Erfassen der Daten je nach Berechnungsmethode entsprechende Rahmenbedingungen erfordert.

5.3. Vergleich mit anderen Arbeiten

Bei der Recherche durchforschte ich folgende Web-Seiten: Osteopathic Research, Pubmed - MEDLINE-Portal und Science Direct. Ich suchte nach Reliabilitätsstudien die manuelle Befundmethoden der Wirbelsäule untersuchten. Im Bereich der Chiropraktoren und Manual-Mediziner (Triggerpoint-Methode) wurde diesbezüglich in den letzten Jahren einiges unternommen. Vorwiegend drei manuelle

Untersuchungsmethoden der Wirbelsäule wurden in diesen Studien auf ihre Intra- und Inter-Rater-Reliabilität untersucht. Dies waren der Befund über Schmerzpunkte (Triggerpunktmethode oder Tenderpointmethode), der Befund über einzelne Wirbelkörper und der Befund über mehrere Segmente. Der Befund der einzelnen Wirbelkörper entspricht dem Inhalt meiner Thesis. Ich stellte mir die Frage, ob es manuelle Befund-Methoden gibt, bei denen eine Reliabilität sei es Intra- oder/und Inter-Rater bereits nachgewiesen werden konnte. Ich stiess auf einige Reviews, welche im Zeitraum von 1965 bis 2008 solche Studien auf deren Qualität und Aussage hin überprüften. Vier davon analysierte ich näher. Dabei interessierten mich vor allem deren Ergebnisse in Bezug auf Reliabilität.

Bei allen vier Reviews kamen die Untersuchenden in etwa zum selben Schluss. Im Zeitraum von 1965-2008 wurden einige hundert Studien publiziert, die sich dem Thema der Reliabilität widmeten. Das Problem fast aller dieser Studien war deren Mangel an Qualität, und so konnten nur sehr wenige Studien (11, 29 und 9) zur Evaluation in die verschiedenen Reviews eingeschlossen werden. Von allen diesen eingeschlossenen Studien waren nur zwei von guter Qualität, der Rest war von mangelhafter bis schlechter Qualität. Mangels besserer Vergleichsmöglichkeiten wurden sie dennoch untersucht. Das Fazit aller Reviews ist aber, dass wegen dem Mangel an Qualität der durchgeführten Studie keine Empfehlungen betreffend die Reliabilität der untersuchten Methoden gemacht werden können. Sie appellieren dafür, dass mehr qualitativ hochstehende Studien durchgeführt werden, damit Empfehlungen herausgegeben werden können. Im Moment gibt es keine nachweislich reliable Methode.

Die Auswertung der oben erwähnten Studien erfolgte mit dem Kappa-Index. Die Berechnungsformel beim Kappa-Index berücksichtigt zwar den Faktor der zufälligen Übereinstimmung, muss aber für die Berechnung von Nominalskalierungen angepasst werden. Wurde diese Formel wegen dem Zufallsfaktor dem Holsti vorgezogen? Was waren die Probleme bei der Durchführung der Studien, dass die Qualität so schlecht ausfiel? Da fast alle Studien über den Qualitätsmangel gestolpert sind, ist es nicht abwegig, nach den Gründen zu fragen die zu diesem Ergebnis führten. Die Wahl des Koeffizienten bestimmt auch das Erfassen der Daten und fordert deshalb entsprechende Rahmenbedingungen. Dem muss man bei der Durchführung der Studie Rechnung tragen. Hat das eventuell die Qualität

beeinträchtigt? Es stellt sich erneut die Frage nach dem geeigneten Berechnungsmodell.

Die Auswertung der Kappa- Werte ergab bei allen Studien dieselben Resultate:

- Für die Befundmethode der einzelnen Wirbelkörper gibt es keine Reliabilität.
- Für die regionale Befundmethode gibt es keine Reliabilität
- Für die Triggerpoint-Methode und Tenderpoint-Methode gibt es eine vage bis akzeptable Reliabilität. Aber da die Qualität der Studien schlecht war und bei erneuter Prüfung die Resultate nicht reproduziert werden konnten, gibt es für diese Methode auch keine Empfehlung für ihre Reliabilität.

Einzig bei den Befund-Methoden mit Schmerzauslösung war eine Reliabilität nachweisbar. Es wurde eine Studie über die Triggerpunkt-Methoden verfasst, welche eine gute Reproduzierbarkeit erzielte. Bei Folgestudien konnten diese Werte jedoch nicht bestätigt werden und deshalb gilt diese Methode im Moment als nicht reproduzierbar. Damit von einer Reproduzierbarkeit gesprochen werden kann, müssen mehrere unabhängige, qualitativ gute Studien die Ergebnisse bestätigen.

Palpatorische Befund-Methoden hatten die Werte „nicht brauchbar“, das heisst Werte unter 0.2.

Die Kappa-Werte können nicht mit den Holsti-Werten verglichen werden. Aber die Aussagen in Bezug auf die Reliabilität dagegen schon. Die Werte meiner Studie deuten auf eine Intra- und Inter-Rater-Reliabilität für den Wirbelsäulenbefund nach der *LMO-Methode* hin. Solche Werte wurden bei den anderen Methoden in ihrer Skalierung nicht erreicht.

Fazit: Es muss ein Modell für manuelle Reliabilitätsprüfungen ausgearbeitet werden, damit die Qualität der Arbeiten steigt und somit auch diesbezüglich Aussagen gemacht werden können. Da die Ergebnisse dieser Studie vielversprechend sind, sollte unbedingt überprüft werden, ob diese reproduzierbar sind. Dies ist nötig, um eine Empfehlung bezüglich der Reproduzierbarkeit dieses Befundes zu geben.

5.3.1. A review of intraexaminer and interexaminer reliability of static spinal palpation: a literature synthesis (2009)

Bei dieser Studie handelt es sich um eine Literaturstudie. Überprüft wurden die Qualität und die Evaluation der Resultate gemachter Intra- und Inter-Rater-Reliabilitäts-Studien über die statische Palpation der Wirbelsäule und der Sacroiliacalgelenke. Es wurden Studien im Zeitraum von 1965-2007 evaluiert. Von anfänglich 343 Studien konnten nur 29 ausgewertet werden. Untersucht wurden drei Typen von Palpation der Wirbelsäule: Palpation von Schmerzpunkten oder Tenderpoints, Palpation mithilfe von Landmarken und Palpation über Knochenstrukturen. Die Reliabilität aller drei Typen ist schlecht.

Die Reproduzierbarkeit bei zwei Studien betreffend die Palpation bei Schmerzauslösung war annehmbar, aber die Qualität dieser Studie dafür schlecht. Statistisch war der Unterschied zu den anderen Studien nicht signifikant.

Zur Berechnung dienten der Kappa-Index und der Übereinstimmungskoeffizient (dieser lag nicht über 50%).

5.3.2. A systematic, critical review of manual palpation for identifying myofascial trigger points: evidence and clinical significance (2008)

Bei dieser Studie wurde die Qualität und klinische Signifikanz der Reproduzierbarkeit der Befunderhebung nach der Methode der Triggerpoints untersucht. Von 1965-2007 wurden verschiedene medizinische Datenbanken durchforstet. Auf Grund der schlechten Evidenz der gefundenen Studien konnten nur 11 berücksichtigt werden. Davon wiesen nur 2 eine gute Evidenz auf und fünf weitere mussten ausgeschlossen werden aufgrund der Ein- und Ausschlusskriterien. Von allen getesteten Punkten wiesen nur drei eine moderate Reliabilität auf. Die übrigen beschreiben eine vage bis schlechte Reliabilität. Die Auswertung erfolgte über den Kappa-Index.

Die methodologische Qualität fast aller Studien dieser Methode war schlecht bis unbrauchbar. Es braucht mehr qualitativ gute Studien, um eine Aussage betreffend die Reliabilität dieser Methode machen zu können.

5.3.3. Reliability of physical examination for diagnosis of myofascial trigger points: a systematic review of the literature (2009)

Bei dieser Studie ging es darum zu untersuchen, ob die Diagnostik von muskuloskelettalen Problemen mit der Methode der Triggerpunktdiagnostik reliabel ist. Auch hier konnten mangels Evidenz nur neun Studien eingebunden werden, wobei keine einzige eine gute Evidenz aufwies. Anhand der Ergebnisse kann diese Methode nicht als reliable Befundmethode zur Untersuchung des Muskuloskelettsystems empfohlen werden. Die Auswertung erfolgte über den Kappa-Index.

5.3.4. Reliability of spinal palpation for diagnosis of back and neck pain: a systematic review of the literature (2004)

Bei dieser Studie werden verschiedene manuelle diagnostische Wirbelsäulenuntersuchungen auf ihre Reliabilität geprüft. Auch hier ist die Evidenz schlecht. Die Resultate werden mit dem Übereinstimmungsindex angegeben. Auch hier müssen mehr Nachweise erbracht werden. Die Reproduzierbarkeit für die Schmerzmethode ist moderat reliabel. Es gibt keine Reproduzierbarkeit für die segmentale palpatorische Diagnostik.

5.4. Vorschläge für anknüpfende Projekte

Bestätigung der Ergebnisse: Die Ergebnisse der Studie sehen vielversprechend aus und sollten weiter untersucht werden. Um eine gesicherte Aussage in Bezug auf die Reliabilität der Befunderhebung der Wirbelsäule nach der *LMO-Methode* machen zu können, müssen diese Resultate reproduziert werden. Erst wenn das der Fall ist, kann von einer Reliabilität gesprochen werden. Beim Durchführen der Studie stellte sich heraus, dass ein Training der Tester vor der Studie nötig gewesen wäre. Also sollte bei einer erneuten Untersuchung dieser Punkt unbedingt beachtet werden.

LMO-Methode: Um in Bezug auf die *LMO-Methode* als ganzheitliche Methode eine Aussage machen zu können, reicht die durchgeführte Studie bei weitem nicht. Es müssen in allen *acht Einheiten* weitere Untersuchungen durchgeführt werden. Hier gibt es viel Arbeit.

Dominante Läsion: Für die Bestimmung der *Dominanten Läsion* sollte das Studiendesign angepasst werden. In meiner Praxis habe ich mit der Bestimmung der

Dominanten Läsion gute Erfahrungen gemacht. Ist das Ermitteln dieser *Läsion* wirklich nur eine Momentaufnahme oder steckt mehr dahinter? Um diese Hypothese zu überprüfen würde ich vorschlagen, ausschliesslich eine Intra-Rater-Reliabilität durchzuführen, bei der jeder Proband nur zwei Mal untersucht wird.

Validitätsüberprüfung des *Provokationstests*: Die Idee dieses Tests beruht darauf, den Wirbelkörper nach vorne zu schieben und so eine Spannung der Gelenks-Kapsel der Facettengelenke zu provozieren. Hier wäre erstens zu überprüfen, ob die Kapsel wirklich gespannt wird und zweitens ob es zu einer Mobilisation der Facettengelenke kommt. Diese Untersuchung erlaubt, Aussagen in Bezug auf die Validität des *Provokationstests* zu machen.

Orthosympathischer Grenzstrang: Einige Probanden hatten deutliche vegetative Reaktionen. In der Osteopathie besteht allgemein (ref) die Hypothese, dass der orthosympathische Grenzstrang durch das Ventralisieren der Wirbelkörper im Bereich der Brustwirbelsäule stimuliert wird. In Bezug auf die hier durchgeführte Studie kann die Hypothese aufgestellt werden, dass es so sein könnte. Es steht aber nur fest, dass vegetative Reaktionen ausgelöst wurden. Es steht nicht fest, welche Struktur diese Symptome ausgelöst hat. Diese Hypothese müsste verifiziert werden.

Berechnungskoeffizienten: Die Wahl des „richtigen“ Berechnungskoeffizienten erwies sich als schwierig. Eine Evaluation dieser Koeffizienten und deren Interpretation für klinische Reliabilitätsstudien fehlen in der Literatur und wären dringend nötig.

5.5. Bedeutung der Ergebnisse für die Osteopathie und Ausblick

Die *LMO-Methode* ist wie jede osteopathische Methode eine manuelle Methode. Das heisst, für den Befund werden ausschliesslich die Hände und deren Sensibilität genutzt. Diese Sensibilität ist sehr individuell und schlecht überprüfbar. Wie bei allen manuellen Methoden ist es deshalb schwierig, die Reproduzierbarkeit eines Befundes wissenschaftlich zu untersuchen und zu bestätigen. Ein weiterer erschwerender Faktor ist die Tatsache, dass Patienten untersucht werden und jeder Patient anders ist. Die wichtigste Voraussetzung für eine Reliabilitätstestung ist aber die Konstanz der Rahmenbedingungen. Aus diesen Gründen ist meiner Meinung nach der Anspruch an die Systematik und Einfachheit der Methode umso grösser.

Ich ging im Vorfeld davon aus, dass die *LMO-Methode* genau diese Kriterien erfüllt. Die Ergebnisse waren so erstaunlich gut, dass wir am Anfang an ihrer Richtigkeit zweifelten. Von der Recherche im Internet wusste ich, wie schwierig es ist, bei manuellen Untersuchungsmethoden der Wirbelsäule eine Reliabilität nachweisen zu können. Die Ergebnisse lassen die Annahme zu, dass die Standardisierung und Einfachheit der Tests den Anforderungen genügen. Für eine gesicherte Aussage zu dieser Reliabilität braucht es Folgestudien, welche die Resultate bestätigen.

Die Osteopathie erhebt den Anspruch, einen sehr individuellen Befund zu erheben. Kein Befund und keine Behandlung gleichen den anderen. Dies macht es schwierig, eine Reproduzierbarkeit des Befundes zu haben. Dasselbe Problem haben andere komplementärmedizinische Methoden wie Homöopathie, Akupunktur und so weiter auch (Melchart, D. et al. 2005). Die Lösung sehe ich auch hierzu in der Anwendung eines standardisierten und einfachen Testverfahrens. Die Methode muss jedoch umfassend und detailliert sein, damit die Individualität des Patienten nicht eingeschränkt wird. Ich sehe in der *LMO-Methode* ein solches Potential. Mit dieser Thesis konnte gezeigt werden, dass die durchgeführte Wirbelsäulenuntersuchung für diese Studie reproduzierbar war. Es lohnt sich, hier weiter zu forschen und zu schauen, ob diese Werte in grösser angelegten Studien reproduziert werden können. Falls dies der Fall ist, macht es Sinn, weitere Befundabschnitte nach ihrer Reliabilität zu untersuchen.

Ich denke dass das Studie-Design akkurat ausgewählt werden muss. Dazu gehört meiner Meinung nach auch die Evaluation der geeigneten statistischen Auswertungsmethode (siehe 5.2). Das Abwägen der Vor- und Nachteile einer Berechnungsmethode ist wichtig, um die Ergebnisse richtig einordnen zu können.

Falls es gelänge, die Reproduzierbarkeit dieses Wirbelsäulenbefundes zu bestätigen, gäbe es in der Osteopathie erstmals eine reproduzierbare manuelle Wirbelsäulen-Befundmethode. Da der Bekanntheitsgrad der *LMO-Methode* am steigen ist und dies auch einhergeht mit der Frage nach wissenschaftlichen Nachweisen, hätte man diesbezüglich etwas in den Händen.

Falls sich in weiteren Studien diese Daten reproduzieren lassen, wäre eines der Gütekriterien erfüllt und man könnte daran anknüpfend eine Validitätsprüfung durchführen.

6. Zusammenfassung

Schlüsselwörter: *Lien Mécanique Ostéopathique*, LMO-Befundmethode, Wirbelsäulen-Befund, *Partielle Totale Läsion*, *Dominante Läsion*, Reproduzierbarkeit, Intra- und Inter-Rater-Reliabilität, Standardisierung hoch genug.

Zielsetzung: Die *Lien Mécanique Ostéopathique (LMO-Methode)* ist eine manuelle Untersuchungs- und Behandlungsmethode entwickelt von Chauffour P. und Prat E. (2003). Sie zeichnet sich durch ihre hohe Standardisierung des Befundablaufs und dessen ausführlichem Protokoll aus. Bei dieser Studie soll getestet werden, ob diese Standardisierung hoch genug ist, um reproduzierbar zu sein. Überprüft wurde die Reproduzierbarkeit des Wirbelsäulen-Befundes nach der *LMO-Methode*.

Forschungsfrage: Sind die durch den Wirbelsäulenbefund nach der *LMO-Methode* ermittelte *Partielle Totale Läsion* (Summe des ermittelten Elastizitätsverlustes) und bestimmte *Dominante Läsion* (Ort mit dem stärksten Elastizitätsverlust) reproduzierbar? Getestet werden sowohl die Intra-Rater- als auch die Inter-Rater-Reliabilität der *Partiellen Totalen Läsion* und der *Dominanten Läsion*.

Methodologie: Methodologische Studie zur Ermittlung der Reliabilität des Wirbelsäulenbefundes nach der *LMO-Methode*. Erstellung einer FMEA (Fehlermöglichkeits- und Einflussanalyse), um im Vorfeld möglichst Fehlerquellen erkennen und beheben zu können. 36 Probanden wurden von je vier studienunabhängigen Testern je zwei Mal untersucht. Der Korrelationskoeffizient wurde mit der Formel von Holsti ermittelt. Zur Bestätigung der Intra-Rater-Reliabilität ist ein Koeffizient von mindestens $r_H:0.75$ gefordert, für die Inter-Rater-Reliabilität $r_H:0.6$.

Ergebnisse: Die Hypothese der Reproduzierbarkeit der Ermittlung der *Partiellen Totalen Läsion* konnte bestätigt werden. Der Intra-Rater-Reliabilitäts-Koeffizient ergab einen Mittelwert von $r_H: 0.794 (\pm 0.098)$. Der Inter-Rater-Reliabilität-Koeffizient ergab einen Mittelwert von $r_H: 0.677 (\pm 0.068)$. Für den Befund der *Dominanten Läsion* konnte die Hypothese in beiden Fällen nicht bestätigt werden. Der Mittelwert des Koeffizienten für die Intra-Rater-Reliabilität betrug $r_H:0.465 (\pm 0.501)$ und für die Inter-Rater-Reliabilität $r_H: 0.181 (\pm 0.191)$

Konklusion: Es gibt eine gute Reliabilität, sowohl Intra-Rater als auch Inter-Rater, für die Ermittlung der *Partiellen Totalen Läsion* bei Wirbelsäulenbefund nach der *LMO-Methode*. Es gibt keine Reproduzierbarkeit für die Bestimmung der *Dominanten Läsion*.

Eine Recherche in Osteopathic Research, Pubmed - MEDLINE-Portal und Science Direct ergab, dass es bisher keine manuelle Befundmethode der Wirbelsäule gibt mit reliablen Ergebnissen.

Die Resultate dieser Studie sind vielversprechend. Es bedarf aber wiederholter Überprüfungen, damit man eine gesicherte, allgemein gültige Aussage bezüglich der Reproduzierbarkeit machen kann.

7. Referenzen

7.1. Abbildungen

Abbildung 1 <i>Provokationstest</i> ganze Wirbelsäule (Quelle: LMO Lehrkörper)	19
Abbildung 2 <i>Provokationstest</i> TH1 (Quelle: LMO Lehrkörper)	19
Abbildung 3 <i>Balancetest</i> BWS (Quelle: LMO-Lehrkörper)	20
Abbildung 4 <i>Recoil</i> , laterales Tibiaplateau nach posterior (Quelle: LMO-Lehrkörper)	21
Abbildung 5 Axiale Druckrichtungen beim <i>Provokationstest</i> (Quelle: LMO-Lehrkörper)	23
Abbildung 6 Ablauf der LMO-Diagnose (Quelle: LMO-Lehrkörper)	23
Abbildung 7 Testung der Wirbelsäule (Quelle: LMO-Lehrkörper)	24
Abbildung 8 <i>Provokationstest</i> (Quelle: LMO-Lehrkörper)	30
Abbildung 9 Auszug einer Auswertung mit PRAM.....	33
Abbildung 10 Beschriftung und Haltung der Probanden.....	39
Abbildung 11 Flowchart Ein- und Ausschluss.....	42
Abbildung 12 Anzahl und prozentuale Verteilung des Geschlechts.....	43
Abbildung 13 Verteilung der Schmerz-Symptomatik in der Stichprobe	44
Abbildung 14 ICC Holsti's Intra-Rater – <i>Partielle Totale Läsion</i>	49
Abbildung 15 ICC Holsti's Inter-Rater – <i>Partielle Totale Läsion</i>	50
Abbildung 16 ICC Holsti's Intra-Rater – <i>Partielle Totale Läsion</i> – Verteilung der Ergebnisse.....	51
Abbildung 17 ICC Holsti's Intra-Rater – <i>Dominante Läsion</i>	52
Abbildung 18 ICC Holsti's Inter-Rater – <i>Dominante Läsion</i>	53
Abbildung 19 ICC Holsti's Intra-Rater – <i>Partielle Totale Läsion</i> = f(Status)	83
Abbildung 20 ICC Holsti's Intra-Rater – <i>Partielle Totale Läsion</i> = f(Gender)	85

7.2. Tabellen

Tabelle 1 Massnahmen für die Anwendung der Auswertung nach Holsti.....	32
Tabelle 2 Zusammenstellung FMEA.....	36
Tabelle 3 Verteilung der Stichprobe nach Alter und Geschlecht.....	43

7.3. Bibliografie

BARKEY,W. (2004): Glossar. : werner.barkey.lu/Uni
BORTZ,J. und DÖRING,N (1995): Forschungsmethoden und Evaluation für Sozialwissenschaftler (2. Aufl.). . : Berlin, Heidelberg: Springer.
BRENNAN, R. L. und PREDIGER, D. J. (1981): Coefficient κ : Some uses, misuses, and alternatives. : Educational and Psychological Measurement. 41, 687–699
CARTER, B. (2003): Methodological issues and complementary therapies: researching intangibles? Complementary Therapies in Nursing and Midwifery, 9(3), 133-139. :
CHAITOW, L. (2002): Neuromuskuläre Techniken in der Manuellen Medizin und Osteopathie. München: Urban&Fischer Verlag München
CHAUFFOUR, P. UND PRAT, E. (2003): Le Lien Mécanique Ostéopathique: Théorie et pratique. Vannes Cedex- France: Sully Éditions
CHAUFFOUR,P.; PRAT,E. und MICHAUD,J. (2008): Die osteopathische <i>Dysfunktion</i> der Arterie. Hippokrates: DO Deutsche Zeitschrift für Osteopathie Thiene, 1/2008, 6. Jahrgang/16-19
CHAUFFOUR,P.; PRAT,E. und MICHAUD,J. (2009): Le Lien Mécanique Ostéopathique: artères et système neuro-végétatif. Vannes Cedex- France: Sully Éditions
DIN EN 60812:2006 (2006): Analysetechniken für die Funktionsfähigkeit von Systemen - Verfahren für die Fehlzustandsart- und -auswirkungsanalyse (FMEA) (IEC 60812:2006); Deutsche Fassung EN 60812:2006. : DIN EN 60812:2006
DORNIEDEN, R (2008): Intra-and Inter-Rater-Reliability of the Test: Anteromedial Position of the Thalus in theThalocrural Joint . Osteopathic Research: WSO (Wiener Schule für Osteopathie)
FRITZSCHE, C. (2008): Randomisierte klinische Studien (RCT), vom Goldstandard zum Sorgenkind. Weblink: H-Blog-Homöopathie & Forschung, 1. September 2008 www.psychophysik.com

FRÜH, W. (2007): Inhaltsanalyse. Theorie und Praxis. Konstanz: UVK/UTB.
GREENHALGH, T. (2003): Einführung in die Evidence-based Medicine, Kritische Beurteilung klinischer Studien als Basis einer rationalen Medizin. Bern: Hans Huber
GRIMES, DA UND SCHULZ, KF (2002): An overview of clinical research: the lay of the land. : Lancet 359 57-61
HANELINE, MT. und YOUNG, M. (2009): A review of intraexaminer and interexaminer reliability of static spinal palpation: a literature synthesis. San Jose, CA USA, Palmer College of Chiropractic West: J Manipulative Physiol Ter. 2009 Jun; 32(5): 379-86
HARTMANN, C (2005): Das Grosse Still-Kompendium, Autobiographie. Forschung und Praxis. Philosophie der Osteopathie. Philosophie und mechanische Prinzipien der Osteopathie. : Jolandos e.K
Hayes, A.F. UND KRIPPENDORFF, K. (2007): (2007): Answering the Call for a Standard Reliability Measure for Coding Data. . : Communication Methods and Measures 1(1),77–89
HIATT, H.und Goldman L (1994): Making medicine more scientific. : Natur 371
HÜSLER, J. UND ZIMMERMANN, H. (2006): Statistische Prinzipien für medizinische Projekte. Hogref AG, Bern: Hans Huber Verlag
ISO (2004): DIN EN 60812, QS-9000. 77–89.:
ISO/TS 16949 (2008): Particular requirements for the application of ISO 9001:2008 for automotive production and relevant service part organizations. : ISO
KIENE, H. (2001): Komplementäre Methodenlehre der klinischen Forschung, Cognition-based Medicine. Berlin, Heidelberg, New York: Springer-Verlag www.ifaemm.de
KIENLE, G. et al. (2003): Evidenzbasierte Medizin: Konkurs der ärztlichen Urteilskraft?. : Dtsch Arztebl 2003; 100(33): A-2142 / B-1784 / C-1688 THEMEN DER ZEIT www.aerzteblatt.de

KOOL, J. und DE BIE, R. (2001): Der Weg zum wissenschaftlichen Arbeiten: ein Einstieg für Physiotherapeuten. Stuttgart New York: Thieme Verlag
KRIPPENDORFF, K (2007): Computing Krippendorff's Alpha-Reliability. : Internet www.asc.upenn.edu
LIEDTKE, D. (2004): Intertester-Reliabilität, Spezifität und Sensitivität des visceralen Tensionstest nach Helsmoortel et al.. Luzern: Swiss Association of Osteopathic Medicine (SAOM)
LIENERT, G.A. und RAATZ, U. (1998): Testaufbau und Testanalyse. München: Beltz Verlag, Psychologie Verlags Union
LOMBARD, M. ; SNYDER-DUCH J. und CAMPANELLA BRACKEN C. (2008): Practical Resources for Assessing and Reporting Intercoder Reliability in Content Analysis Research Projects. Review: Temple University, Philadelphia, PA, USA
LUCAS, L. et al. (2009): Reliability of physical examination for diagnosis of myofascial trigger points: a systematic review of the literature . School of Biomedical and Health Sciences, University of Western Sydney, Auckland, New Zealand: Clin J Pain, 2008 Jan;25(1):80-9
MAYRING, P. (2002): Einführung in die Qualitative Sozialforschung. Weinheim und Basel: Beltz Verlag
MAYRING, P. (2005): Qualitative Inhaltsanalyse, Grundlagen und Techniken. Weinheim und Basel: Beltz Verlag 2005, 183ff
MELCHART, D. (2003): Universitäre Forschung in Naturheilkunde und Komplementärmedizin – Perspektiven und Standortbestimmung. München: Zentrum für naturheilkundliche Forschung II. Med. Klinik und Poliklinik der Technischen Universität München content.karger.com
MELCHART, D. et al. (2005): Programm Evaluation Komplementärmedizin(PEK), Schlussbericht. Bern: Bundesamt für Gesundheit www.bag.admin.ch
MERTEN, K (2004): Inhaltsanalyse. Einführung in Theorie, Methode und Praxis 2., verbesserte Auflage. Opladen: Westdeutscher Verlag

MERTEN,K (1995): Inhaltsanalyse: Einführung in Theorie, Methode und Praxis,2., verb. Aufl.. Opladen : Westdeutscher Verlag
MYBURGH, C. et al. (2008): A systematic, critical review of manual palpation for identifying myofascial trigger points: evidence and clinical significance. Institute of Sports and Clinical Biomechanics, University of Southern Denmark, : Arch Phys Med Rehabil, 2008 Jun;89(6): 1169-76
NEUENDORF, K. A. (2009): PRAM (Version 0.4.5) und seine Bedienungsanleitung . : www.geocities.com
NEUENDORF, K. A. (2002): The content analysis guidebook. : Thousand Oaks, CA: Sage
PORTNEY, LG UND WATKINS, MP (2000): Foundations of clinical research; applications to practice, 2. Aufl.. New Jersey: Prentice Hall Health
Psychische Diagnostik (s.a.): . : www.verhaltenstherapeutin.com
PUTZ, R. und PABST, R. (2007): Sobotta, Anatomie des Menschen, Der komplette Atlas in einem Band. München: Urban&Fischer Verlag München
RASPE, H. (s.a.): Konzept und Methode der Evidenz-basierten Medizin: Besonderheiten, Stärken, Grenzen, Schwächen und Kritik. Campus Lübeck: Institut für Sozialmedizin Universitätsklinikum Schleswig-Holstein
RICHTER, R. und HEBGEN, E. (2008): Triggerpunkte und Muskelfunktionsketten in der Osteopathie und Manuellen Medizin. 2. aktualisierte Auflage. : Thieme
RÖSSEL-BRETSCHNEIDER, A. & MONTEIRO-FERREIRA, J. und THUILLIER L. (2002): Untersuchung der Reproduzierbarkeit osteopathischer Tests am Beispiel des Beckens. Wien: WSO (Wiener Schule für Osteopathie)
RÖSSLER, P. (2005): Inhaltsanalyse. Konstanz: UVK/UTB.
RUSELL, B. (2004): Mechanical Link: Fundamental Principles, Theory, and Practice Following an Osteopathic Approach. . : Physical Therapy, Book Review, Jan 1, 2004
SACKETT, D. (1996): Evidence based medicine: what it is and what it isn't. :

McMaster University, BMJ-Artikel 1996 www.ebm-netzwerk.de
SCHUMACHER, M. und SCHULGEN, G. (2008): Methodik klinischer Studien, Methodische Grundlagen der Planung, Durchführung und Auswertung. Berlin, Heidelberg: Springer-Verlag
SEFFINGER, MA. et al (2004): Reliability of spinal palpation for diagnosis of back and neck pain: a systematic review of the literature. Department of Osteopathic Manipulative Medicine, College of Osteopathic Medicine, Pomona CA USA: Spine (Phila Pa 1976), 2004 Oct 1;29(19):413-25
SMOKVINA, J. (2007): Seminar Gesundheit und Krankheit: Recht und Medizin im Dialog; Komplementärmedizin im Recht. Zürich: Rechtswissenschaftliches Institut www.rwi.uzh.ch
SOMMERFELD, P. (2006): Methodologie, Kleine und hoffentlich kurzweilige Einführung in den Sinn und Unsinn wissenschaftlichen Arbeitens auf der Grundlage quantitativer Methoden für OsteopathInnen. Wien: Wiener Schule für Osteopathie
SOMODY-NEPLAZ, I. (2007): Der Mechanical Link. Elsevier: Osteopathische Medizin 8(1):4-9
STATISTA (2009): Definition Reliabilität. : de.statista.com
VON EYE, A. (2006): An Alternative to Cohen's κ . : European Psycholog. 11, 12-24
WINDELER, J. ("s.a."): Evidenzklassierungen . : www.ebm-netzwerk.de
"N.N." (2008): Evidenzbeurteilung. Weblink: Cochrain www.cochrane.de
"N.N." (2008): Grundlagen der Evidenzbasierten Medizin. Weblink: Deutsches Netzwerk Evidenzbasierte Medizin e. V www.ebm-netzwerk.de
"N.N." (2004): Reporting Guidelines. : Thieme www.thieme-connect.de
"N.N." (2008): Der Wirksamkeitsnachweis in der Medizin. Freiburg i. Brsg.: Institut für angew. Erkenntnistheorie und med. Methodologie (IFAEMM) www.ifaemm.com
"N.N." (2009): Berechnung von Reliabilitätskoeffizienten. : app.gwv-fachverlage.de

8. Anhang

8.1. Dokumente

8.1.1. Probandeninformation

osteoteam.ch, C. Hafen-Bardella, Lyssachstr. 9, 3400 Burgdorf

Sehr geehrte Probandin, sehr geehrter Proband

Vielen Dank, dass Sie sich bereit erklären an meiner Grundlagenstudie „**Intra- und Interrater Reliabilität der Befundtechnik der Wirbelsäule nach Lien Mécanique Ostéopathique (LMO)** teilzunehmen.

Bei dieser Grundlagenforschung geht es darum zu überprüfen, ob verschiedene Osteopathen, welche dieselbe manuelle Untersuchungsmethode zur Beurteilung der Beweglichkeit der einzelnen Wirbelkörper anwenden, zum gleichen Schluss kommen. Diese Überprüfung wird Interrater Reliabilität genannt.

Bei der Intrarater Reliabilität geht es darum, dass derselbe Osteopath den Befund beim gleichen Probanden zwei Mal (mit Unterbruch dazwischen) durchführt. Die Fragestellung in diesem Fall ist, ob diese zwei Befunderhebungen übereinstimmen.

Der Nutzen dieser Studie liegt darin zu sehen, ob man mit der Befunderhebung nach LMO eine manuelle Untersuchungsmethode hat die reproduzierbar ist. Das heisst dass das Befundergebnis nicht abhängig ist von der Interpretation und dem Spürsinn des einzelnen Osteopathen, sondern dass bei korrekt angewandter Methode jeder Osteopath zum selben Schluss kommt.

In der Manuellen Befunderhebung der Beweglichkeit der Wirbelkörper gibt es in der ganzen Klassischen und Alternativen Medizinwelt noch keine Methode, die dieses Kriterium erfüllt. Dies gilt sowohl für die Befunderhebung durch verschiedene Therapeuten, als auch für die Wiederholung der Befunde durch dieselbe Person.

Befundmethode:

Jeder Wirbelkörper wird mittels leichtem Druck auf den Dornfortsatz auf seine Elastizität hin überprüft. Jeder Wirbelkörper der nicht elastisch ist wird festgehalten. Diese Wirbelkörper werden untereinander verglichen und der dominante unter ihnen bestimmt. Die Befundmethode ist schmerzlos. Es findet keine Behandlung statt.

Behandlungsablauf

Die Untersuchung findet am 16. 05 in der Praxis osteoteam.ch an der Lyssachstrasse 9 in Burgdorf gemäss Plan statt.

Vor der Untersuchung werden alle Probandinnen und Probanden von Claudia Hafen für die Untersuchung vorbereitet. Das beinhaltet die Erhebung einiger Daten und die Beschriftung der Wirbelkörper mittels Klebezettel welche nach der Befunderhebung wieder entfernt werden.

Jede Probandin, jeder Proband wird insgesamt 8 Mal von 4 Osteopathen untersucht. Bei den Osteopathen handelt es sich um Kollegen die seit Jahren ausschliesslich nach dieser Methode ihre Befunde erheben und teilweise diese Methode auch unterrichten.

Der Zeitaufwand beträgt ca. 1h.

Als Gegenleistung für ihren Zeitaufwand hat jede Probandin und jeder Proband Anrecht auf eine osteopathische Behandlung bei Claudia Hafen. Falls eine Behandlung gewünscht ist kann diese individuell vereinbart werden.

Information Probanden

osteoteam.ch, C. Hafen-Bardella, Lyssachstr. 9, 3400 Burgdorf

Risiken

Die Teilnahme an der Studie birgt für die Probandinnen und Probanden keine Risiken.
Es werden keine Wirkungen erwartet, da es sich bei der Studie ausschliesslich um eine Befunderhebung handelt und nicht um eine Behandlung.
Es ergeben sich für die Probandinnen und Probanden keine Vor- beziehungsweise Nachteile durch die Teilnahme an der Studie

Beilagen

- Schriftliche Einverständniserklärung zur Teilnahme an der Studie
- Zeitplan

Damit verbleibe ich mit freundlichen Grüssen

C. Hafen-Bardella

Information Probanden

8.1.2. Schriftliche Einverständniserklärung

Kantonale Ethikkommission Bern (KEK), Postfach 56, 3010 Bern

Schriftliche Einverständniserklärung des Patienten zur Teilnahme an einer Grundlagenstudie

- Bitte lesen Sie dieses Formular sorgfältig durch.
- Bitte fragen Sie, wenn Sie etwas nicht verstehen oder wissen möchten.

Nummer der Studie	
Titel der Studie:	Intra- und Interrater Reliabilität der Befundtechnik der Wirbelsäule nach Lien Mécanique Ostéopathique (LMO)
Ort der Studie:	Burgdorf, Kanton Bern
Prüfende Osteopathen/-innen Name und Vorname:	Claudia Hafen-Bardella / Angelika van der Kraats / Regula Studer Benassi / Diane Mottier / Dominique Moreau
Versuchsperson Name und Vorname:	
Geburtsdatum:	Geschlecht:
Gesetzlicher Vertreter der Versuchsperson (falls er unmündig, entmündigt oder nicht urteilsfähig ist): Name und Vorname:	
Evtl. Verwandtschaftsgrad der Versuchsperson	
Zeuge (falls erforderlich): Name und Vorname:	

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

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

Ich hatte genügend Zeit, um meine Entscheidung zu treffen.

Ich bin darüber informiert, dass eine Versicherung allfällige Schäden deckt, falls solche im Rahmen der Studie auftreten.

Ich bin einverstanden, dass die Studienleitung zu Prüf- und Kontrollzwecken in meine Originaldaten Einsicht nehmen darf, jedoch unter strikter Einhaltung der Vertraulichkeit.

Ich nehme an dieser Studie freiwillig teil. Ich kann jederzeit und ohne Angabe von Gründen meine Zustimmung zur Teilnahme widerrufen.

Ort, Datum	Unterschrift der Versuchsperson
Burgdorf, 16. Mai 2009	
Ort, Datum	ev. Unterschrift des gesetzlichen Vertreters/ Zeugen
Burgdorf, 16. Mai 2009	
Ort, Datum	Unterschrift Claudia Hafen, Studienleitung
Burgdorf, 16. Mai 2009	

8.1.3. Stamblatt der Probanden mit Ein- und Ausschlusskriterien

Geändert	15.07.2009 14:48
Name	
Geburtsdatum	
Geschlecht	<input type="radio"/> weiblich <input type="radio"/> männlich
Rückenschmerzen	<input type="radio"/> Nie <input type="radio"/> selten <input type="radio"/> oft <input type="radio"/> ständig
Sind diese Rückenschmerzen ausstrahlend?	<input type="radio"/> Ja <input type="radio"/> Teilweise <input type="radio"/> Nein
Bei Ausstrahlungen: wo?	
Treten die Rückenschmerzen auf	<input type="radio"/> nur nach Belastung <input type="radio"/> auch ohne Belastung <input type="radio"/> gar nicht
Wann treten die Schmerzen auf	<input type="radio"/> Nachts <input type="radio"/> Morgens <input type="radio"/> Abends <input type="radio"/> Irgendwann <input type="radio"/> Nie
Lassen sich die Schmerzen auslösen durch	<input type="radio"/> Bewegung <input type="radio"/> Druck <input type="radio"/> Sonstiges <input type="radio"/> Gar nicht
Wie ist ihre Belastung im Alltag	<input type="radio"/> gering <input type="radio"/> mittel <input type="radio"/> gross
Welchen Beruf üben Sie aus	
Sind zurzeit die Schmerzen	<input type="radio"/> akut <input type="radio"/> subakut <input type="radio"/> still
Welche Therapien haben Sie gehabt?	keine
Welche Operationen hatten Sie	
Welche Medikamente nehmen Sie regelmässig ein	
Welche Medikamente nehmen Sie bei Bedarf ein	
Bemerkungen	
Editoren	
Geändert von	Hafen Claudia

8.1.4. Befundblatt

Proband		Therapeut	
Befund			
C0	<input type="radio"/> Links <input checked="" type="radio"/> Negativ <input type="radio"/> Rechts	TH 7	<input type="radio"/> Positiv <input checked="" type="radio"/> Negativ
C1	<input type="radio"/> Links <input checked="" type="radio"/> Negativ <input type="radio"/> Rechts	TH 8	<input type="radio"/> Positiv <input checked="" type="radio"/> Negativ
C2	<input type="radio"/> Positiv <input checked="" type="radio"/> Negativ	TH 9	<input type="radio"/> Positiv <input checked="" type="radio"/> Negativ
C3	<input type="radio"/> Positiv <input checked="" type="radio"/> Negativ	TH 10	<input type="radio"/> Positiv <input checked="" type="radio"/> Negativ
C4	<input type="radio"/> Positiv <input checked="" type="radio"/> Negativ	TH 11	<input type="radio"/> Positiv <input checked="" type="radio"/> Negativ
C5	<input type="radio"/> Positiv <input checked="" type="radio"/> Negativ	TH 12	<input type="radio"/> Positiv <input checked="" type="radio"/> Negativ
C6	<input type="radio"/> Positiv <input checked="" type="radio"/> Negativ	L 1	<input type="radio"/> Positiv <input checked="" type="radio"/> Negativ
C7	<input type="radio"/> Positiv <input checked="" type="radio"/> Negativ	L 2	<input type="radio"/> Positiv <input checked="" type="radio"/> Negativ
TH 1	<input type="radio"/> Positiv <input checked="" type="radio"/> Negativ	L 3	<input type="radio"/> Positiv <input checked="" type="radio"/> Negativ
TH 2	<input type="radio"/> Positiv <input checked="" type="radio"/> Negativ	L 4	<input type="radio"/> Positiv <input checked="" type="radio"/> Negativ
TH 3	<input type="radio"/> Positiv <input checked="" type="radio"/> Negativ	L 5	<input type="radio"/> Positiv <input checked="" type="radio"/> Negativ
TH 4	<input type="radio"/> Positiv <input checked="" type="radio"/> Negativ	Coccygis	<input type="radio"/> Positiv <input checked="" type="radio"/> Negativ
TH 5	<input type="radio"/> Positiv <input checked="" type="radio"/> Negativ	Ilium (posteriore Seite) Anterior	<input type="radio"/> Links <input checked="" type="radio"/> Negativ <input type="radio"/> Rechts
TH 6	<input type="radio"/> Positiv <input checked="" type="radio"/> Negativ	Ilium (posteriore Seite) Posterior	<input type="radio"/> Links <input checked="" type="radio"/> Negativ <input type="radio"/> Rechts
Dominante	<auswählen>	Bemerkungen	
Wirbelsäule	Erstellt 15.07.2009 14:51	Geändert	15.07.2009 14:51
Autor	Hafen Claudia	Geändert von	Hafen Claudia
Editoren			

8.2. Weitere Angaben zur Statistik

Folgende drei Koeffizienten kommen für die Berechnung in Frage:

- Übereinstimmungskoeffizient ($r_{\bar{U}}$)
- Reliabilitätskoeffizient nach Holsti (r_H)
- Krippendorff's α (α)

Zusammenfassung Koeffizient	Intra				
	1	2	3	4	Avg
Übereinstimmung %	0.824	0.751	0.829	0.773	0.794
Holsti's	0.824	0.751	0.829	0.773	0.794
Scott's Pi	0.549	0.282	0.548	0.556	0.484
Cohen's Kappa	0.551	0.413	0.554	0.559	0.519
Spearman Rho	0.555	0.412	0.562	0.549	0.520
Pearson Korrelation	0.547	0.399	0.559	0.567	0.518
Lin Konkordanz	0.538	0.389	0.544	0.560	0.508
Krippendorff's Alpha	0.565	0.429	0.549	0.555	0.524

Zusammenfassung Koeffizient	Inter (Mittel aller Paare)		
	Run1	Run2	Avg
Übereinstimmung %	0.677	0.677	0.677
Holsti's	0.677	0.677	0.677
Scott's Pi	0.232	0.244	0.238
Cohen's Kappa	0.255	0.270	0.263
Spearman Rho	0.272	0.290	0.281
Pearson Korrelation	0.272	0.290	0.281
Lin Konkordanz	0.255	0.270	0.263
Krippendorff's Alpha	0.332	0.293	0.313

8.3. Einfluss Status – Symptomatisch – Asymptomatisch

Es konnten keine signifikanten Unterschiede der Resultate zwischen erstem und zweitem Durchgang in Abhängigkeit des Status festgestellt werden.

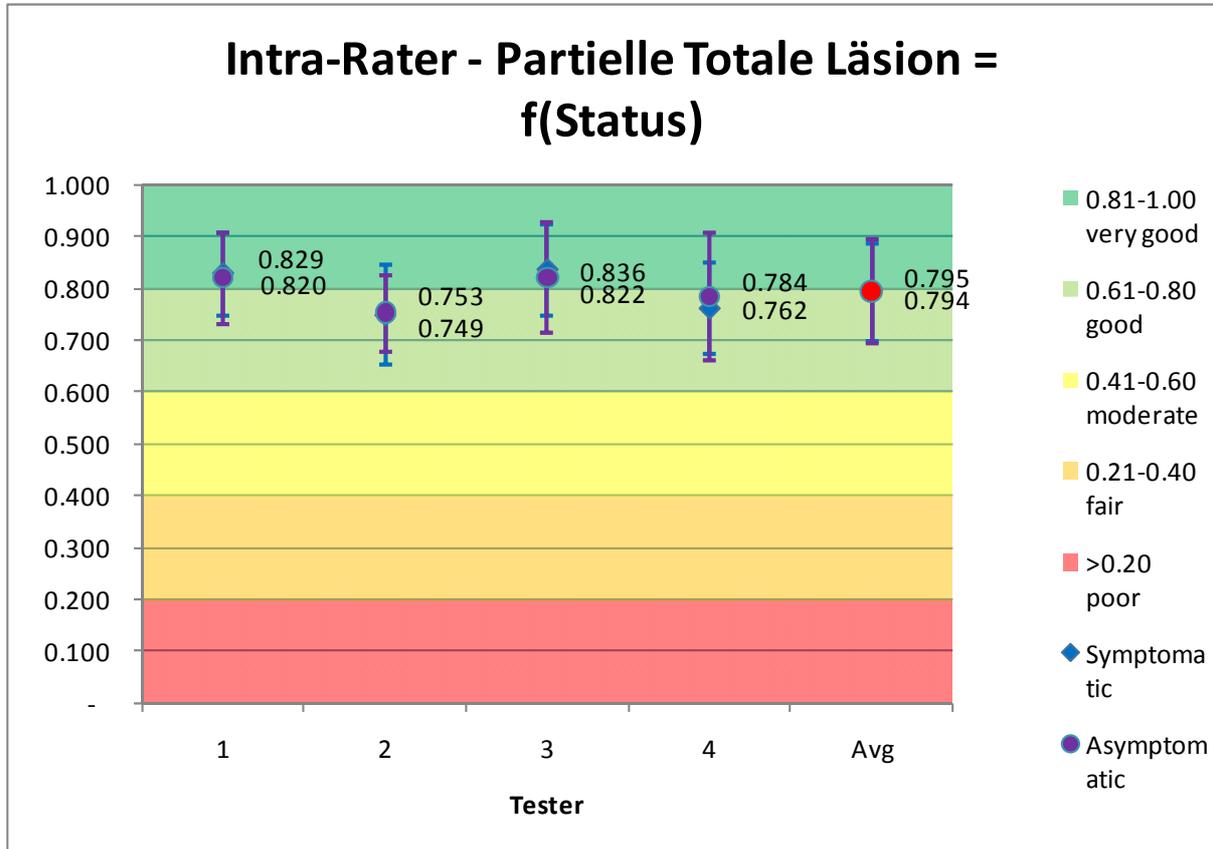
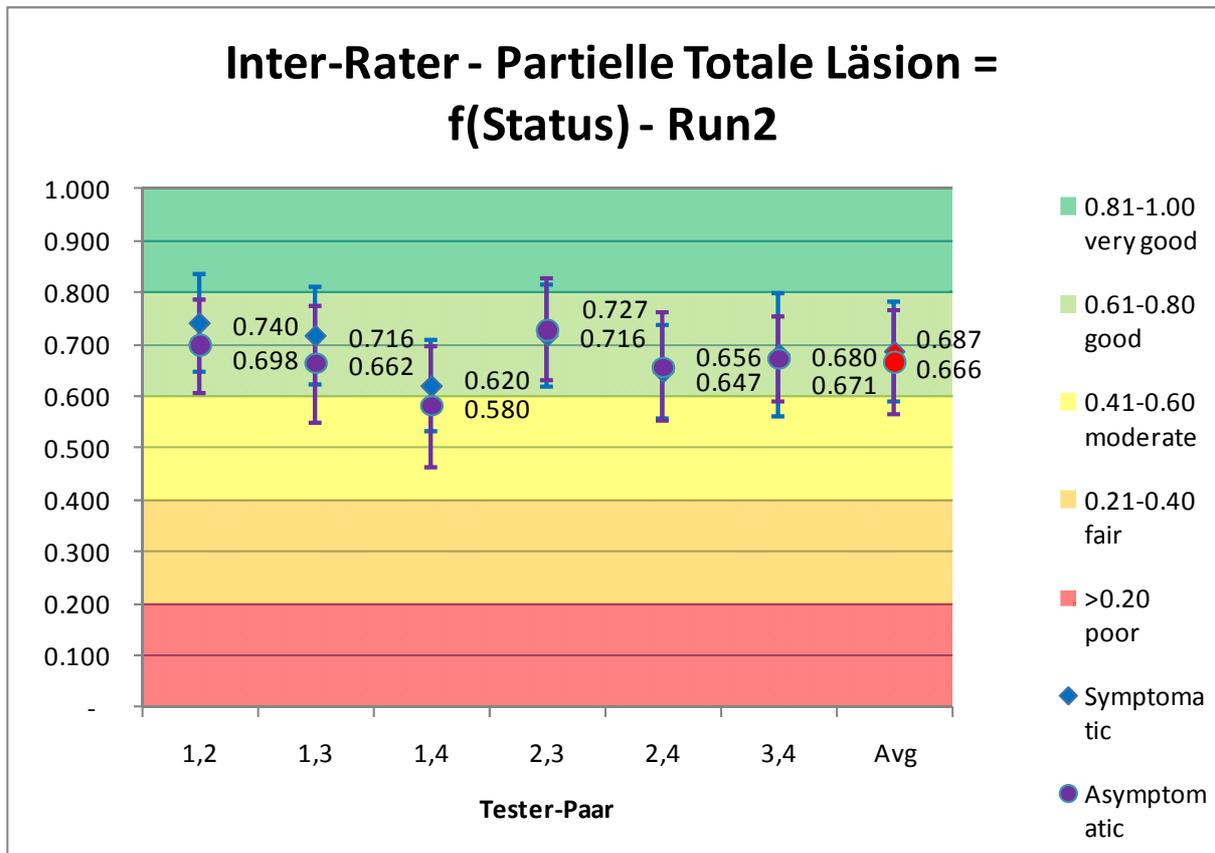
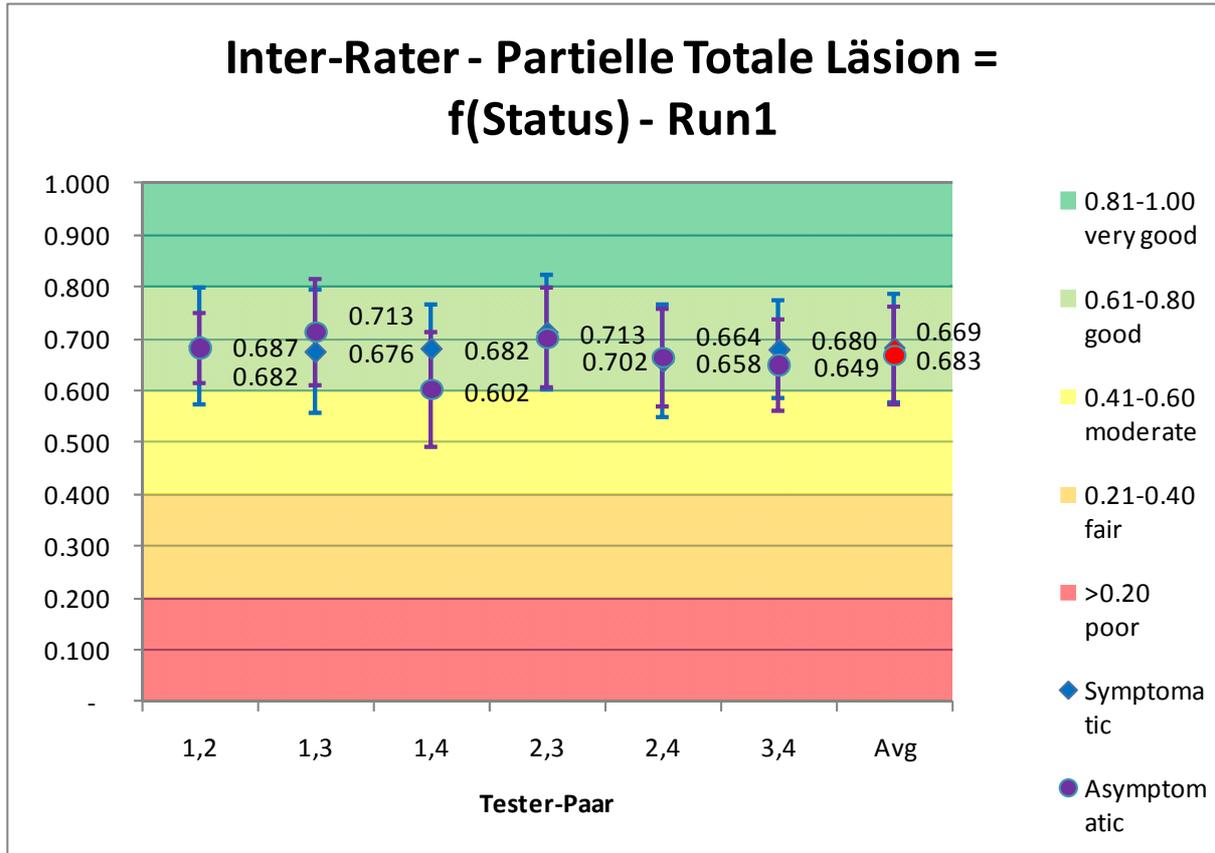


Abbildung 19 ICC Holsti's Intra-Rater – Partielle Totale Läsion = f(Status)

Avg = Arithmetisches Mittel der Ergebnisse der vier Tester

ICC = Reliabilitäts-Koeffizient nach Holsti (arithmetisches Mittel der Einzelergebnisse beider Durchgänge eines Testers)

Die ermittelten Koeffizienten unterscheiden sich unwesentlich zwischen Intra- und Inter-Rater.



8.4. Einfluss Geschlecht

Auch der Einfluss des Geschlechts auf das Resultat zeigte keine wesentlichen Unterschiede.

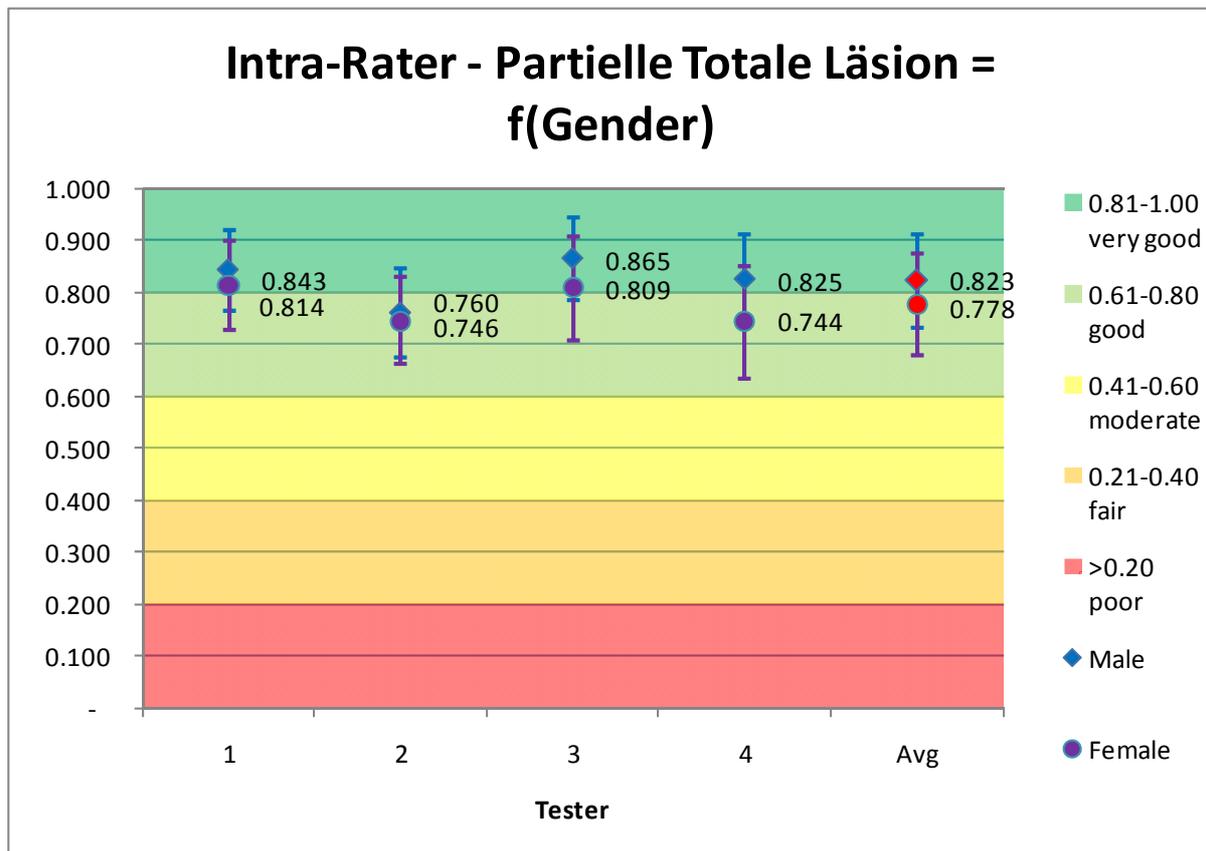
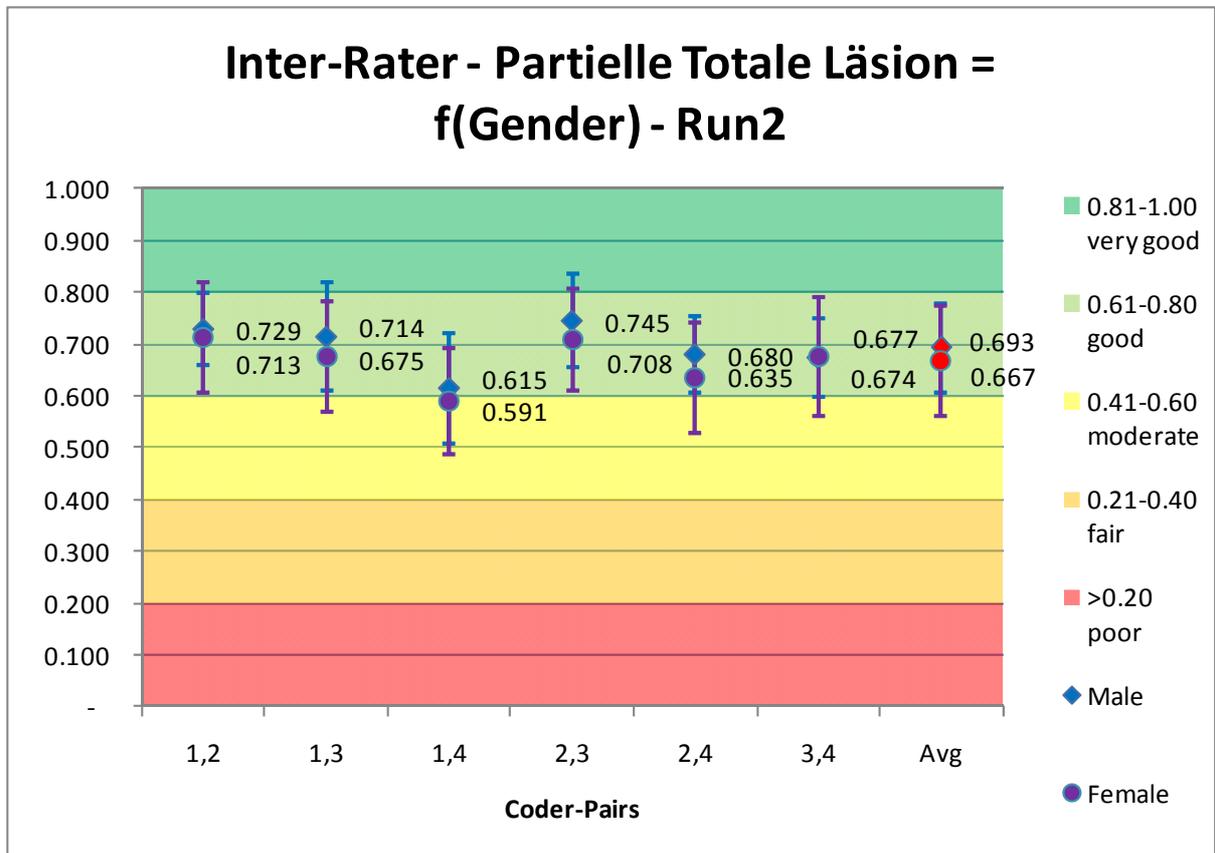
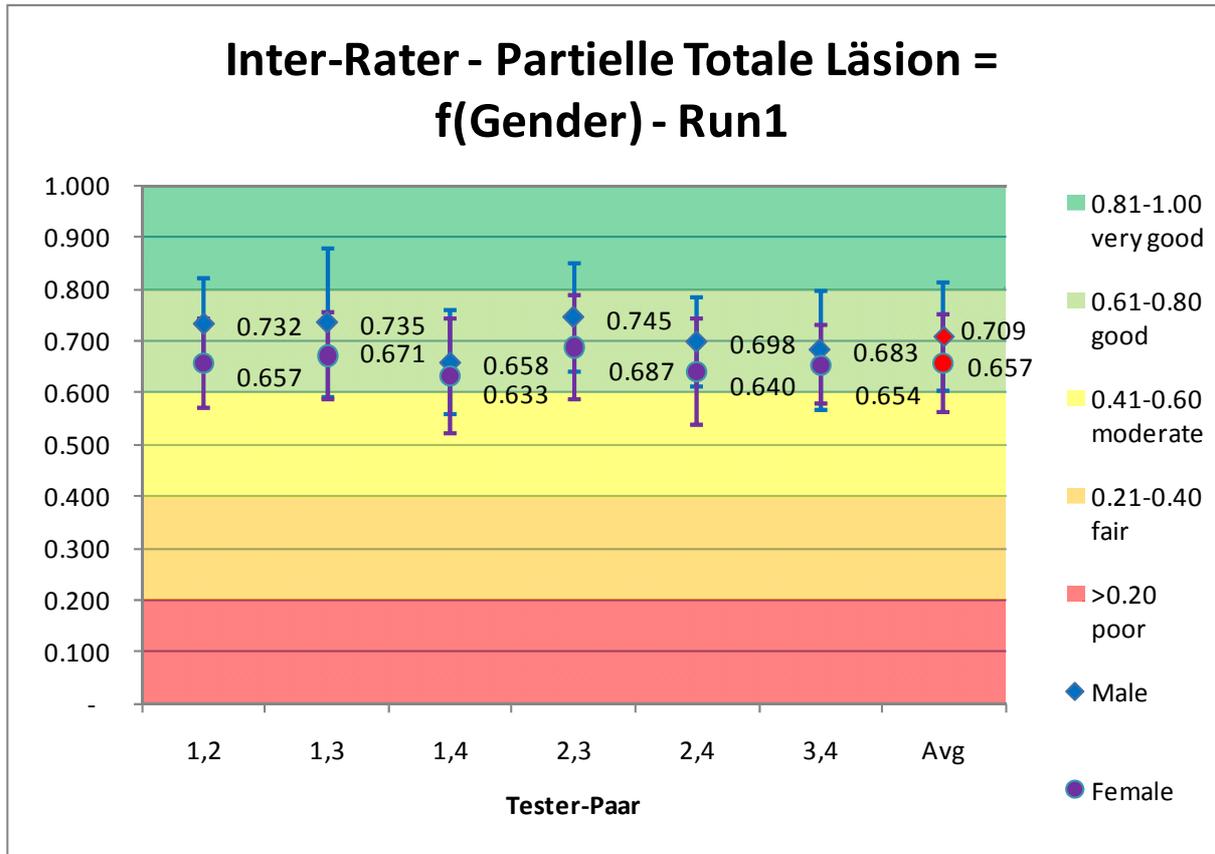


Abbildung 20 ICC Holsti's Intra-Rater – Partielle Totale Läsion = f(Gender)

Avg = Arithmetisches Mittel der Ergebnisse der vier Tester

ICC = Reliabilitäts-Koeffizient nach Holsti (arithmetisches Mittel der Einzelergebnisse beider Durchgänge eines Testers)

Die ermittelten Koeffizienten unterscheiden sich unwesentlich zwischen Intra- und Inter-Rater.



Reliabilitätsstudie des Befunds nach LMO

Reliabilitätsstudie über die Befunderhebung der
Wirbelsäule nach der Methode der
Lien Mécanique Ostéopathique

Summary (Short version of the Thesis in English)

Master Thesis zur Erlangung des Grades
„Master of Science“ in Osteopathie

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

niedergelegt
an der **Wiener Schule für Osteopathie**

von ***Claudia Hafen-Bardella***

Burgdorf, November 2009

Betreut von *Claudia Gamsjäger und Friederike Kaiser*
Übersetzt von *Jack Hafen*

Abstract

Key words: Reproducibility, LMO finding method, *Lien Mécanique Ostéopathique*, spine findings, partial total lesion, dominant lesion, intra- and inter-rater reliability.

Objective: The *Lien Mécanique Ostéopathique* (LMO method) is a manual examination and treatment method developed by P. Chauffour and E. Prat. With the spine findings the partial total lesion (sum of all positive tests found) and the dominant lesion (maximum positive test) are determined.

Research Question: Are the partial total lesion and the dominant lesion reproducible when they are determined by the spine findings according to the LMO method?

Methodology: A methodological study of the intra-rater and the inter-rater-reliability of the spine findings according to the LMO method. The size of the sample has been fixed at 30 test persons at least. After applying the inclusion and the exclusion criteria 36 test persons could be included in the study. The homogeneity of the test persons could be guaranteed. Four testers, independent of this study, do the findings twice with every test person. The testers had no insight into the files of the test persons, neither before nor during the study. The correlation coefficient was determined with the formula of Holsti. For the confirmation of the intra-rater-reliability a coefficient of equal/greater than: r_H 0.75 is required, for the inter-rater-reliability r_H 0.6.

Results: The hypothesis of the reproducibility of the determination of the partial total lesion was confirmed. The Intra-rater-reliability-coefficient showed an average of r_H : 0.794 (± 0.098). The Inter-rater-reliability-coefficient showed an average of r_H : 0.677 (± 0.068). For the findings of the dominant lesion, the hypothesis in both cases could not be confirmed. The mean average of the coefficient for the Intra-rater-reliability was r_H : 0.465 (± 0.501) and for the Inter-rater-reliability r_H : 0.181 (± 0.191).

Conclusion: There is a good reliability, for both the Intra-rater as well as the Inter-rater, for the determination of the partial total lesion with the spine findings according to the LMO method. There is no reproducibility for the determination of the dominant lesion.

Content

1. Introduction	4
1.1. Problem definition.....	4
1.2. Relevance of the thesis for the general public/patients	4
1.3. Relevance of the thesis for the osteopathy.....	4
2. Question / Hypotheses.....	6
2.1. Hpotheses for the partial total lesion	6
2.2. Hypotheses for the dominant lesion	6
3. Bases	7
3.1. Definition of the LMO method	7
3.2. Reliability	9
4. Methodology	10
4.1. FMEA (Failure mode and effects analysis).....	11
4.2. Procedure of the testing	13
5. Results	14
5.1. Determination of the partial total lesion	14
5.2. Normal distribution of the results	16
5.3. Determination of the dominants.....	17
5.4. Side results.....	17
6. Discussion.....	18
7. Continuing studies	22
8. Synthesis, Outlook	23
9. References.....	25
9.1. Bibliography.....	25
9.2. Pictures	26
10. Appendix.....	27
10.1. Normal distribution of the results.....	27
10.2. Determination of the dominants	28

1. Introduction

The demand for complementary medicine methods has continually increased in recent years. Accordingly the number of patients who made use of them increased significantly (Carter, 2003). Critics asked for proofs of efficacy, safety and efficiency of these methods.

1.1. Problem definition

In my practice I have been working since 2005 exclusively by the method of Lien Mécanique Ostéopathique (LMO) developed by Paul Chauffour and Eric Prat (2003) (see 3.1)

This method (Chauffour, Pratt; 2003) is characterized by detailed and systematized findings (approx. 350 tests). The development and further development of the method by the authors has been purely empirical so far. Since I use LMO every day, whether as osteopath or as assistant, my desire for a scientific basis to this method grew more and more. This prompted me to make this matter the theme of my master thesis and to examine the reproducibility = reliability of the spine findings by the LMO-method. The reliability is one of the criteria for an empirical research (Mayring 2005), and according to Mayring (2005) the reliability of a testing method is a criterion for its quality. If the reliability of a method has been found, this is a characteristic to examine the validity of the method.

1.2. Relevance of the thesis for the general public/patients

The reliability of osteopathic findings by the LMO method can be determined by this thesis. In my opinion it is important that manual findings are adjusted, assessed and conveyed. The osteopath has to know how well his present findings can be reproduced, so as to draw conclusions that are generally valid. This is to protect the patients from false information, based on overvalued manual findings.

Avoiding inappropriate treatment by false findings.

1.3. Relevance of the thesis for the osteopathy

The reliability and quality (Mayring, 2005) of major findings of the spine by the LMO method can be determined. If the null hypothesis is correct, the result is not to be exaggerated. If the alternative hypothesis is correct, the findings by this method win in quality and should be further investigated.

Should the standardization of the method be high enough, the thesis serves for the further development of a promising findings model. There is a chance for the first time to have a reproducible manual method. If the reliability cannot be proved, the standardization could be continued.

If the reliability of the findings method of the spine according to LMO can be proved, it would be a sign of quality for these partial findings. The conveyance and learning of this findings method will get a new status. The teaching of LMO is based on a standardized precise instruction, by which the desired anatomical structures are found and which shows on what level and in which direction these structures must be tested. If there is a reproducibility of findings it may also be considered a quality control for the accuracy of the teaching of this method.

The results allow further studies: According to Mayring (2005: 185) it only makes sense to investigate the validity of the method if its reliability has previously been studied.

2. Question / Hypotheses

The question arises whether the standardization in the clinical examination of the spine by the LMO method is high enough to prove the reproducibility.

Using an Intra and an Inter-Rater-reliability-review, I would like to investigate whether the findings by the LMO method are reproducible. Since the findings by the LMO method are bipartite partial total lesion (see 3.1) and dominant lesion (see 3.1), we have the following four scenarios:

2.1. Hypotheses for the partial total lesion

Null hypothesis 1.1. (Intra-Rater): The findings of the investigation of the partial total lesion of the spine by the LMO method is accidental.

Alternate hypothesis: 1.1. (Intra Rater): The findings of the investigation of the partial total lesion of the spine by the LMO method in repeated implementation by the same investigator is predominantly equal (reproducible).

Null hypothesis 1.2. (Inter-Rater): The findings of the investigation of the partial total lesion of the spine by the LMO method is dependent on the investigator.

Alternate hypothesis 1.2. (Inter-Rater): The findings of the investigation of the partial total lesion of the spine by the LMO method is predominantly equal (reproducible), independent on the investigator.

2.2. Hypotheses for the dominant lesion

Null hypothesis 2.1. (Intra-Rater): The findings of the determination of the dominant lesion of the spine by the LMO method are accidental.

Alternate hypothesis: 2.1. (Intra-Rater): The findings of the determination of the dominant lesion of the spine by the LMO method in repeated implementation by the same investigator is predominantly equal (reproducible).

Null hypothesis 2.2. (Inter-Rater): The findings of the determination of the dominant lesion of the spine by the LMO method are dependent on the investigator.

Alternate hypothesis 2.2. (Inter-Rater): The findings of the determination of the dominant lesion of the spine by the LMO method are predominantly equal (reproducible), independent on the investigator.

3. Bases

The emergence of the LMO method (Somodyn-Neplaz, 2007) goes back on the year 1978 and was designed and developed by Paul Chauffour. 1984 Eric Prat joined him. During the last 20 years they expanded together the method of LMO. The LMO method claims to be an independent diagnosis and treatment model.

The idea of Chauffour (2003: 39ff) is based on the understanding that dysfunctions in the sense of LMO develop through the loss of elasticity in the connective tissue and fascias. Based on this he examines the body for losses of elasticity.

3.1. Definition of the LMO method

Here are some LMO specific definitions (Somodyn-Neplaz, 2007):

- **Provocationstest**

This is a test to verify the elasticity of the tissue. The connecting tissue is set under tension (=provocation) through targeted, three-dimensional pull or pressure. If no deformation in the sense of LMO results, no elasticity exists and the test is considered positive (lesion exists).

- **Lesion = dysfunction:** positive provocations test
- **Total lesion:** Sum of all positive provocations tests found in the testing of all eight functional units. Normally, all eight units are tested in the LMO method.

- **Partial total lesion:** Sum of all positive provocations tests in the implementation of a partial finding. It is in one or in more units. This is used for emergency treatments as e.g. immediately after a distortion-traumata.
- **Inhibitions test/ Balance test:** The LMO uses the model that the lesions are linked together. Based on this the inhibitions test according to LMO was developed. It serves the hierarchy of the lesions found through the



Picture 1 Provocation Test TH1 (Ref. Teachers group LMO)

provocations test. The strongest such lesion in the LMO model is called the primary lesion.

- **Primary lesion:** The strongest lesion that is found in the balance test in the hierarchy of the total lesion (the positive tests of all eight functional units). This is usually determined in each treatment.
- **Dominant lesion:** The strongest lesion that is determined by the balance test in the hierarchy of the partial total lesion (the positive tests of the tested functional units). This shortened approach is used for emergency treatments.
- **The eight units according to LMO:** The concept of LMO claims, if possible, to test all structures in the body that means to find the total lesion called such according to LMO. Therefore the body is divided in eight functional units.

Spine testing: In my thesis, the reproducibility of the spine findings was examined according to the LMO method. For this reason an axial pressure will be applied in the course of the acantha as a provocations test according to the anatomical incidents. The intention in this test is to put under tension both the upper spine joint as well as the lower spine joint. It is assessed whether an elasticity is felt or not. This assessment is subjective.



Picture 2 Spine Testing with the provocation test (Ref. Teachers group LMO)

The clinical examination is highly standardized. It is systematic and always done according to the same protocol. Therefore it is very well suited to verify the reliability of the results. The LMO diagnosis is made in two steps. The first step is the identification of the partial total lesion using a provocations test and in the second step the positive provocations tests are hierarchized using the inhibitions test so as to determine the dominant lesion (Somodyn-Neplaz, 2007).

3.2. Reliability

The reliability (the reproducibility of a method) is one of the three main quality criteria and is a measure of the usefulness or the reliability of clinical tests or measurement methods (Barkley, 2004).

Inter-Rater-reliability: Mayring (2005: 185): Inter-Rater-reliability in the empirical research (including psychology, sociology, epidemiology, ...) the extent of conformity (= concordances) of the appraisal results by different observers (= "raters" or "coders"). A test person is examined by two or more osteopaths. The results should be equal for an Inter-Rater-reliability. So it can be shown to what extent the results are independent from the individual osteopath. The reliability is a measure of the quality of the method for measuring a certain variable.

Intra-Rater-reliability: Mayring (2005: 185): In contrast to the above, the Intra-Rater-reliability shows the conformity of the appraisal results in repeated assessments by the same observer. By this it can be shown to what extent the results are reproducible and do not depend on the current situation.

4. Methodology

This thesis is a methodologic study according to Portney and Watkins (2000). A reliability testing will be made. The Intra-Rater-reliability and the Inter-Rater-reliability in the findings of the spine of the LMO method for determining the partial total lesion and the dominant lesion are investigated (see 3.1). For this purpose four osteopaths and 36 test persons participate in this study. Each osteopath examines each test person two times. The results from the different osteopaths are compared, as well as the results of the two investigations by the individual osteopath. The osteopaths must have at least three years' experience in this technique. They have no insight into the files of the test persons neither before nor during the testing. The data are dictated and fed into the PC by auxiliary staff. Between the first and second finding the osteopath has to make three other investigations. Each finding takes five minutes as a maximum.

Inclusion criteria: unspecific back pains, symptomatic or asymptomatic, with or without radiations in the lumbar and buttocks area (only sensibility disturbances permitted).

Exclusion criteria: radicular symptoms, radicular signs such as reduced reflexes and loss of strength in arms and legs, acute back pains of any kind, acute inflammation of any kind, fever. Missing findings data.

Sample sizes: The literature indicates 30 to 50 codings per category as a minimum sample size for a reliability study. 200 to 300 are deemed desirable (Früh, 1998: 189)

Calculation coefficient: The evaluation of the data is carried out with the Holsti's coefficient. The pairwise matches of the collected findings are calculated.

Confirmation of the hypothesis: For a confirmation of the hypothesis the reliability coefficient for the Intra-Rater-reliability must be higher than r_H : 0.75 and for the Inter-Rater-reliability higher than r_H : 0.6.

4.1. FMEA (Failure mode and effects analysis)

An analysis was made in advance in order to find and avoid possible sources of errors.

Error/Effect	Cause	Action
Not standardized name and localisation of the vertebra	Error of palpations, differing points of reference, manual setting (interpretation) of the vertebra	Preliminary marking of the acantha by an independent person
	The inscription moves when the test person alters his position	Define the position of the test persons and instruct the test persons. The test persons remain for all testings at the same place, the tester switches the area, the tester checks the position before the testing
	Wrong application of the inscription (approach the acantha)	Instruction of the tester how to use the inscription. Precise definition of the spot on the acanthi where the sticker has been placed

Error/Effect	Cause	Action
Not uniform implementation of the tests	Differing handling	The method has to be used by the tester for three years at least
	Differing pressure during the provocations test	
	Lack of experience	
Influencing the tester	Hearing what the predecessor dictates	Separate premises for testing No access to the patients files allowed
	Access to the patients files	
Impact of the testing on the musculoskeletal system	Mobilization of the spine through the ventralisation	Per vertebral body only one testing per round
	Long testing phase	
	Vegetative reactions by ventralisation	
Wrong fed data into the PC	Confusion when entering data	Feedback by the typist
	Misunderstanding between the tester and the person feeding data into the PC	Repetition of the vertebra to be tested

4.2. Procedure of the testing

36 test persons were examined twice in a day by four osteopaths in sequence. Were the partial total lesion and the dominant lesion of the spine (see 3.1) were determined with every test person. The data are complete from all tests.



Picture 3 Testing of the spine

The testing was done in separate rooms, in order to prevent the mutual influence of the testers. In order to prevent access to documents by the testers, they had to dictate the data to auxiliaries on the PC. The data fed into the PC were repeated in order to minimize the errors. All the points of the errors and impact analysis (see 4.1) were considered, and it proved helpful. The testing could be done without problems. After the clinical examination it became apparent that some patients had vegetative reactions such as nausea, dizziness... Since the frequent testing of each vertebra (eight times) caused the risk of vegetative reactions, the testers were instructed to make the test with every vertebra and every passage only once. They didn't pay heed.

5. Results

The results of the spine findings by the LMO method included two aspects: Finding the partial total lesion and determining the dominant lesion. The main findings of the investigation are as follows:

The calculations of the reliability coefficient according to Holsti **confirmed** the hypotheses 1.1 and 1.2. for the determination of the partial total lesion.

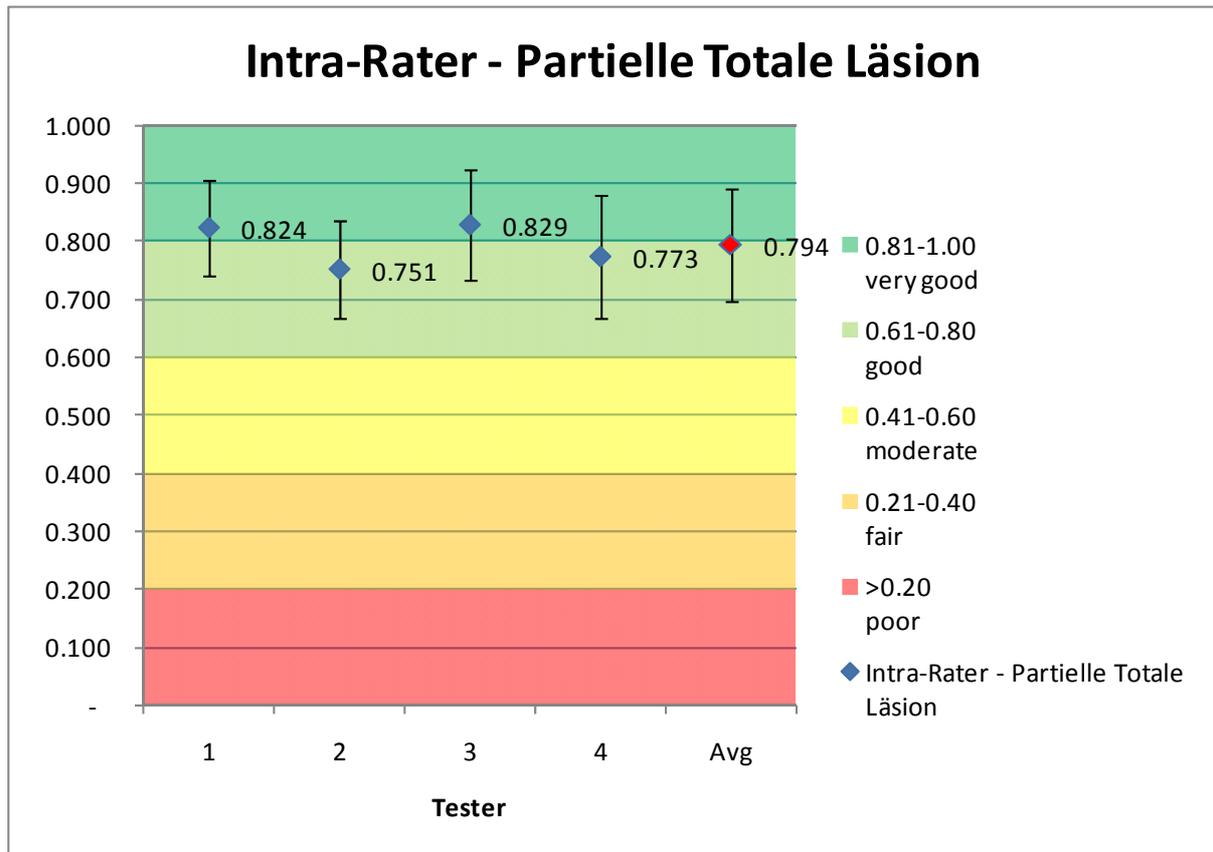
- The hypothesis 1.1.: The reproducibility of the partial total lesion of the Intra-Rater-reliability was confirmed, with an average of all tests of r_H : 0,794 ($\pm 0,098$)
- The hypothesis 1.2.: The reproducibility of the partial total lesion of the Inter-Rater-reliability, with an average of all tests from r_H : 0,677 ($\pm 0,068$) was confirmed.

The calculations of the reliability coefficient according to Holsti **did not confirm** the hypotheses 2.1. and 2.2. for the determination of the dominant lesion.

- The hypothesis 2.1.: The reproducibility of the dominant lesion of the Intra-Rater-reliability, with an average of all tests from r_H : 0,465 ($\pm 0,501$) was not confirmed.
- The hypothesis 2.2.: The reproducibility of the dominant lesion of the Inter-Rater-reliability, with an average of all tests from r_H : 0,167 ($\pm 0,193$) was not confirmed.

5.1. Determination of the partial total lesion

The results of the Intra-Rater-reliability for the determination of the partial total lesion for all four testers are in the area of good and very good. Two testers, whose average value is in the field above r_H : 0.8, two testers are slightly below. The average Intra-Rater-reliability is with 0.79 ($\pm 0,098$) just below the area very well. The small standard deviation shows that 2/3 of all values of the entirety are probably in the area of good and very good (Hüsler/Zimmermann, 2006: 58).



Picture 4 ICC Holsti's Intra-Rater – Partial total lesion

Avg = arithmetic average of the results of all coders

ICC = Reliability-coefficient according to Holsti (arithmetic average of the individual results of both runs of one coder)

The results of the Inter-Rater-reliability for the determination of the partial total lesion are for all four testers in the area well. The average Inter-Rater-reliability is 0.68 ($\pm 0,068$). The small standard deviation shows that 2/3 of all values of the entirety will likely be in the area well (Hüsler/Zimmermann, 2006:).

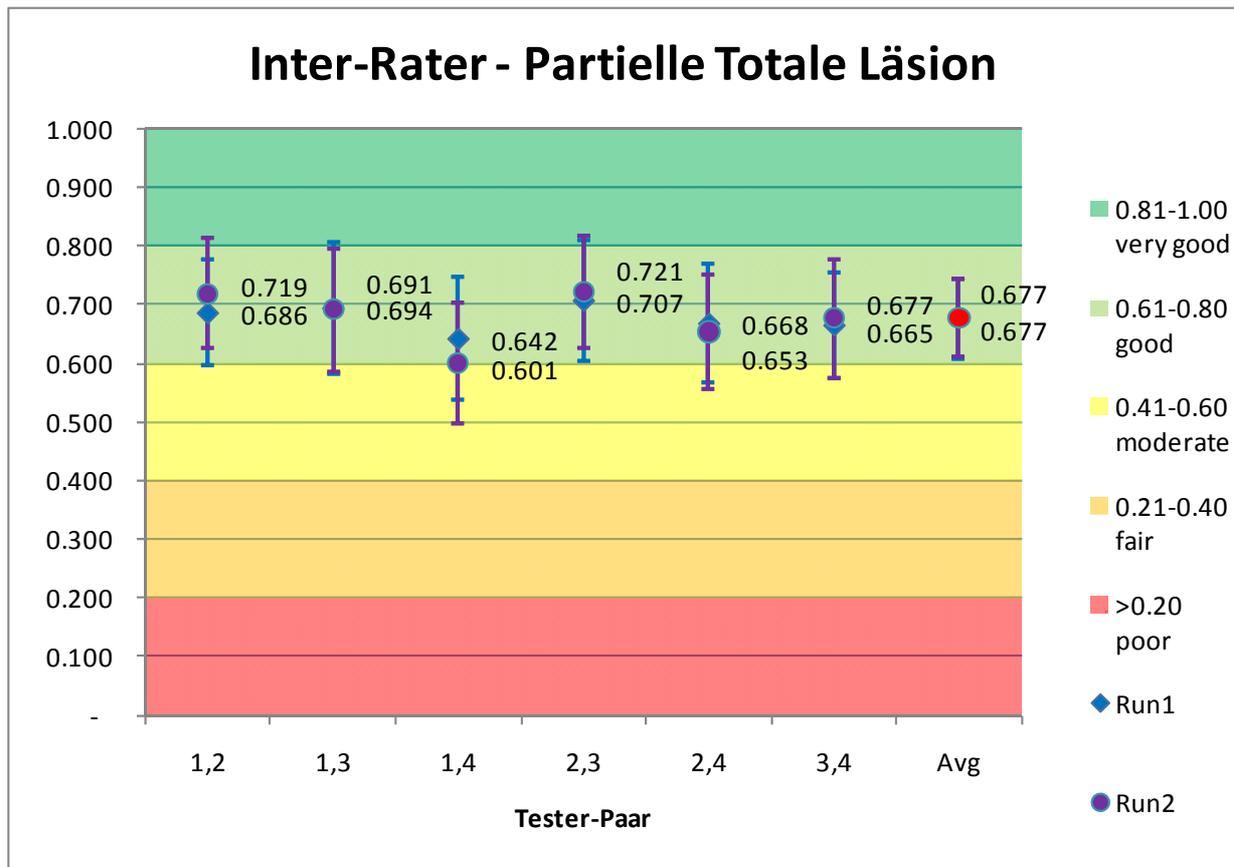


Abbildung 1 ICC Holsti's Inter-Rater – Partielle Totale Läsion

Avg = arithmetic average of the results of all six coder-pairs (for both runs)

ICC = Reliability-coefficient according to Holsti (arithmetic average of the individual results of two testers in the same run. Two results for every tester-pair)

5.2. Normal distribution of the results

The optical assessment of the histograms shows an almost normally distributed sample. On the recommendation of the statistician a wider investigation (such as statistical tests of the obliqueness) was waived, since the range of averages ($AVG \pm SDEV$) with these results does not cover more than two classes. So 2/3 of all values cover two classes at the most. The accuracy of the calculation is in any case higher than the uncertainty of the characterization of the results in themselves (see diagrams in appendix 10.1).

5.3. Determination of the dominants

The average results of the Intra-Rater-reliability for the determination of the dominant lesion of all four testers are located in the area moderate. For two testers the average is in the area r_H : 0.6, for two testers in the area r_H : 0.3.

The average Intra-Rater-reliability is with r_H : 0.47 ($\pm 0,501$) also in the area moderate. The wide standard deviation however shows that 2/3 of the values of the entirety are probably distributed in all classes and are therefore not evidential (Hüsler/Zimmermann, 2006:58) (see also appendix 10.2).

The average results of the Inter-Rater-reliability for the determination of the dominant lesion for all four testers are located in the area useless.

The average Inter-Rater-reliability is with r_H : 0.18 ($\pm 0,191$) also in the area useless. The great standard deviation shows that 2/3 of all values of the entirety are probably in the two lowest classes, and therefore are not evidential (Hüsler/Zimmermann, 2006:58) (see also appendix 10.2).

5.4. Side results

The results in comparison between the first and the second findings, between male and female and between symptomatic and asymptomatic test persons do not differ (Hafen, 2009: 87f)

6. Discussion

Intra and Inter-Rater-reliability of the partial dominant lesion: The reproducibility of the spine findings seems to be well with the Intra-Rater-reliability with an average value of r_H : 0.8. The experts argue about the assessment of such coefficients (Hafen, 2009: 32ff), but for all specialists, a value of 0.8 is well (even if the method by Holsti does not take into account the factor of accidental matches). The value reached in the study sounds promising and therefore should be further explored, especially because such values are rarely seen in manual methods (Hafen, 2009: 64ff). This study is a first step, a first approach, to the review of the reproducibility of the spine findings by the LMO method. Now further steps would have to be taken to check if these results are confirmed and whether other findings respectively part of findings are reproducible.

The number of test persons is sufficient to make a statement. But it is on the lower limit. Since the results were good, it would be worthwhile to carry out the study in a broader context, 2-300 test persons (Hüsler/Zimmermann, 2006) (Hafen, 2009: 45ff).

The method claims to be exact and structured. This standardization is apparently sufficient for an Intra-Rater reproducibility of the spine findings.

The reproducibility of the spine findings with several testers is still acceptable with a value of r_H : 0.7. As said in the statistical part (Hafen, 2009: 32ff), it can be argued as to when a value is evident. But 0.7 is clearly more than the accidental number 0.5. So you can read from this value that there is at least a trend reproducibility. The number 0.7 is to be handled with care, since in the formula by Holsti the factor of accidental match is not taken into account.

The question is interesting why there is a difference between the Intra-Rater and the Inter-Rater-reliability. The following factors could have influenced this: not recognized sources of errors (testers were insufficiently trained), testers did not fully comply with the rules, in a Inter-Rater-reliability the standardization of the method is insufficient or

barely sufficient to compensate for the subjective feeling of the testers in interpreting elasticity.

Handling and standardization of testing: In most studies one of the widely debated factor of errors is the lack of standards in the procedure of findings and the complexity in the assessment of the testing. Just these problems are mentioned in the master thesis of Dornieden (2008). Both the Inter-Rater as well as the Intra-Rater-reliability revealed no viable reproducibility, even when the testers had applied and practiced the method for a long time. Quite the opposite: The results showed that the Intra-Rater-reliability of testers fresh from the school was rather better. The different handling of the tests and the lack of standardization were mentioned as possible sources of errors. An important factor in the reliability test of a measurement method is that the measurement is done in the same way whenever it is possible (Hafen, 2009: 28). In our case this would be the implementation of the tests. The most uniform and precise handling of a test is an important factor for the reproducibility of a method. The testers in this study work daily with the LMO method and are experienced. For each patient the testing of the spine is made regardless of his troubles. The testers are also members of a LMO training group who has been meeting for two years in intervals of two months. In these meetings the handling and the procedure of the findings are practiced and freshend up. In addition, two of the members of this group belong to the team of teachers of Chauffour and Prat and therefore are often engaged with the conveyance of the precise implementation of the tests according to the LMO method. It is assumed that the testers can boast of an unusually high degree of conformable handling. This could be an important factor for the good reliability of the study.

The LMO finding claims to be highly standardized. I believe two issues are relevant. Firstly, the application and implementation of the test are very simple. It does not require complex three-dimensional settings to be implemented and is either positive or negative. This simplicity leaves little room for personal creativity. The simpler a test is the easier is the learning of the same handling by various testers.

Secondly, the procedure of the testing is always done by the same protocol and is repeated with every clinical examination. This could be another factor for the good results of this study.

In this study not only the positive tests were scored as being identical but also the negative tests because they were deliberately coded negative. This resulted in a total of 25 tests per patient and passage. When we verified whether the results were correctly identified, because they looked too well, we found that they would have looked very different if we had included the positive tests only. Most studies include only the positive tests. Perhaps the negative ones should also be included in the calculations provided because they are deliberately tested as negative.

Influence of the sensitivity of the tester to the result: In the Intra-Rater-reliability the findings are made several times by the same tester with the same test person. In this kind of testing the element of the tester's subjective sensation is not significant because it is always the same tester who decides. By contrast, the testing at the Inter-Rater-reliability is done by various testers whose sensibility is different. I tried in advance to reduce the influence of this factor with the measure, that every osteopath has worked at least three years with this method. It proved to be not enough. Apparently a deeper tester training is required. The FMEA analysis should be completed by this point.

The reproducibility of the dominant lesion: The reproducibility of the dominant lesion in the findings of the spine showed useful results neither with the intra nor with the inter-Rater-reliability. The result could not be reproduced, not even by the same tester. To judge by the results, the balance test seems to be a snapshot. It is a question whether the framing conditions for this test were too unsteady (influence by the number of tests) or whether the balance test is not sufficiently standardized. The LMO method calls for the primary lesion. Whether this is possible, is questionable under the circumstances and needs further investigations (Hafen, 2009: 61).

Does the testing of the spine joints change the framing conditions? In order to give a statement regarding the framing conditions, I compared the average coefficients of the first passage with those of the second passage. If the eight tests mobilized the spine joints, we might expect that the framing conditions would change. The results of the first and second round would be likely to be different. This was not the case. So it can be assumed that by the testing the spine joints are not or

inadequately mobilised in order to change the framing conditions. An assumption that the testing provokes a treatment is questionable with this result.

It is an open question whether the testing, because of the possible mobilization of the spine joints had an influence on the framing conditions in relation to the determination of the dominant lesion. To determine this influence, an Intra-Rater-reliability study would have to be carried out. In this study the number of tests would have to be limited to a minimum (only two tests per vertebral body).

The test persons showed vegetativ reactions (Hafen, 2009: 48). In diagnosing the problem, the question arose whether the ventralization of the vertebral body, by the testing, stimulate the sympathetic chain. Whether the vegetative reactions were really caused by the stimulation of the vegetative sympathetic chain through the ventralization of the vertebral body, cannot be answered conclusively.

7. Continuing studies

Confirmation of the results: The results of the study look promising and should be examined further. The results must be reproduced so as to permit a secured statement regarding the reliability of the findings of the spine by the LMO method.

LMO method: Examinations in all 8 units are required in order to make statements about the LMO method as an all-embracing method. There is much work to be done. Since the results are promising, the effort would be worthwhile because so far no proven reliable manual findings method for the spine exists.

Dominant lesion: The study design should be adapted for the determination of the dominant lesion. In order to re-examine this hypothesis I would propose to carry out an Intra-Rater-reliability, investigating every test person only twice.

Validity examination of the provocation test: As a further study on reliability with statements about a possible mobilization of spine joint and the stimulation of the sympathetic chain.

Calculation coefficients: Choosing the "right" calculation coefficient proved difficult. An evaluation of these coefficients and their interpretation of clinical reliability studies are almost entirely missing in the literature and would be urgently needed.

8. Synthesis, Outlook

The LMO method is, as any osteopathic method, a manual method. This means that for the findings one's hands and their sensitivity are used exclusively. This sensitivity is very individual and hard to verify. As with all manual methods, it is therefore difficult to investigate scientifically and to confirm the reproducibility of a finding. Another complicating factor is the fact that patients are examined, and each patient is different. But the most important prerequisite for a reliability testing is the consistency of the framing conditions. For these reasons, I think the requirements for the systematics and the simplicity of the method are all the greater. In the beginning I proceeded on the assumption that the LMO method met exactly these criteria. The results were so astonishingly good that at the beginning we were in doubt about their authenticity. From my research in the Internet I knew how difficult it is to prove a reliability in a manual investigation method of the spine. The results admit the assumption that the standardization and simplicity of the tests fulfil the requirements. For a secure statement about this reliability, we need studies which confirm the results.

If additional studies reproduced this data one of the quality criteria would be met and we could go on from this to carry out a validity test.

Osteopathy claims to reach very individual findings. No finding and no treatment are equal to another. This makes it difficult to reach a reproducibility of a finding. The same problems occur with other complementary medicine methods such as homeopathy, acupuncture etc. (Melchart, D. et al. 2005). The solution I see is the application of a standardized and simple testing. However the method must be comprehensive and detailed, so that the individuality of the patients is not restricted. I see that potential in the LMO method. With this thesis it has been shown that the spine testing has been reproducible for this study. It is worthwhile to continue the research and to see whether these values can be reproduced in larger-scale studies. If this is the case, it makes sense to continue examining finding segments for their reliability.

I think that the study design has to be selected accurately. This includes in my opinion the search for the appropriate statistical evaluation method (Hafen, 2009: 63). The balance between the pros and cons of a calculation method is important for the correct ranging of the results.

If we were able to confirm the reproducibility of these spine findings there would be for the first time in the osteopathy a reproducible manual finding method for the spine. Since the reputation of the LMO method is rising and this is associated with the question of scientific evidence we would have something tangible in the hands.

9. References

9.1. Bibliography

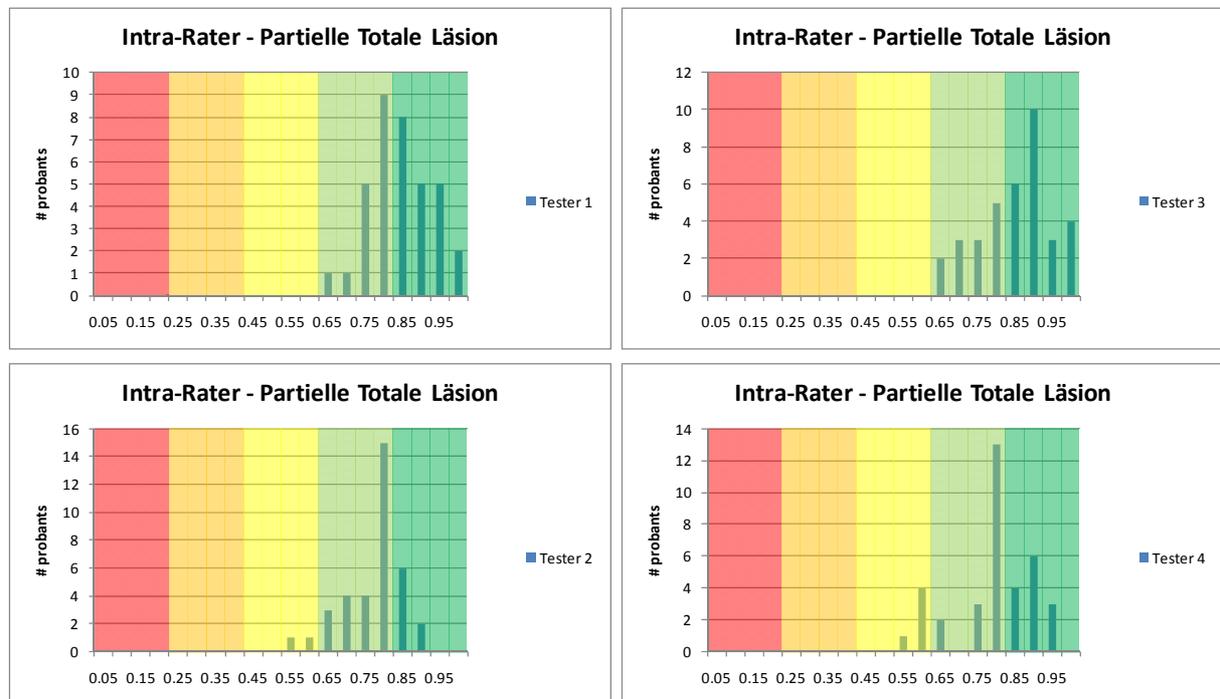
BARKEY,W. (2004): Glossar. : werner.barkey.lu/Uni
CARTER, B. (2003): Methodological issues and complementary therapies: researching intangibles? Complementary Therapies in Nursing and Midwifery, 9(3), 133-139. :
CHAUFFOUR, P. UND PRAT, E. (2003): Le Lien Mécanique Ostéopathique: Théorie et pratique. Vannes Cedex- France: Sully Éditions
DORNIEDEN, R (2008): Intra-and Inter-Rater-Reliability of the Test: Anteromedial Position of the Talus in theThalocrural Joint . Osteopathic Research: WSO (Wiener Schule für Osteopathie)
FRÜH, W. (2007): Inhaltsanalyse. Theorie und Praxis. Konstanz: UVK/UTB.
HÜSLER, J. UND ZIMMERMANN, H. (2006): Statistische Prinzipien fürmedizinische Projekte. Hogref AG, Bern: Hans Huber Verlag
ISO (2004): DIN EN 60812, QS-9000. 77–89.:
MAYRING, P. (2002): Einführung in die Qualitative Sozialforschung. Weinheim und Basel: Beltz Verlag
MELCHART, D. et al. (2005): Programm Evaluation Komplementärmedizin(PEK), Schlussbericht. Bern: Bundesamt für Gesundheit www.bag.admin.ch
PUTZ, R. und PABST, R. (2007): Sobotta, Anatomie des Menschen, Der komplette Atlas in einem Band. München: Urban&Fischer Verlag München
RÖSSLER, P. (2005): Inhaltsanalyse. Konstanz: UVK/UTB.
SOMODY-NEPLAZ, I. (2007): Der Mechanical Link. Elsevier: Osteopathische Medizin 8(1):4-9

9.2. Pictures

Picture 1 Provocation Test TH1 (Ref. Teachers group LMO)	7
Picture 2 Spine Testing with the provocation test (Ref. Teachers group LMO)	8
Picture 3 Testing of the spine	13
Picture 4 ICC Holsti's Intra-Rater – Partial total lesion	15
Picture 5 Holsti's Intra-Rater - Partial total lesion - Distribution of the results	27
Picture 6 Holsti's Intra-Rater - Dominant lesion	28
Picture 7 Holsti's Inter-Rater - Dominant lesion	29

10. Appendix

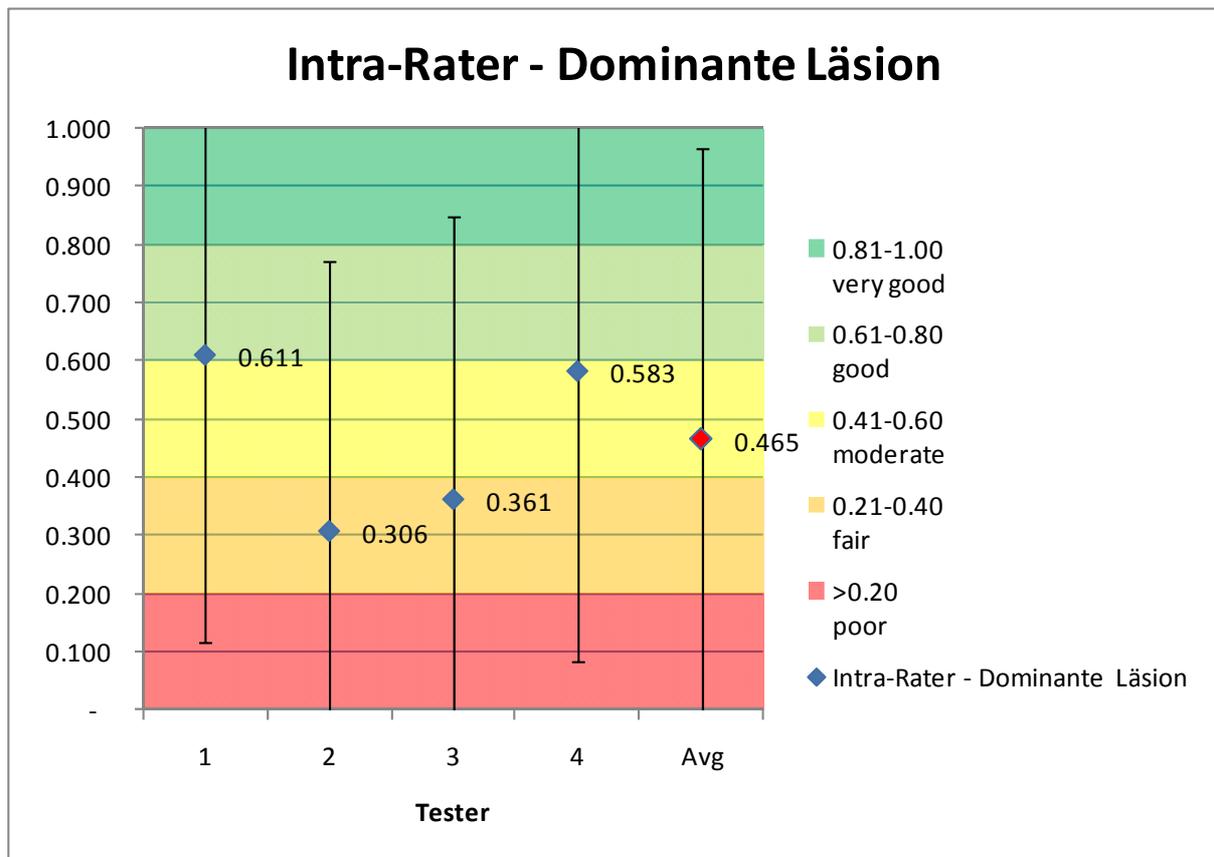
10.1. Normal distribution of the results



Picture 5 Holsti's Intra-Rater - Partial total lesion - Distribution of the results

The optical assessment of the histograms shows an almost normally distributed sample. On the recommendation of the statistician a wider investigation (such as statistical tests of the obliqueness) was waived, since the range of averages ($AVG \pm SDEV$) with these results does not cover more than two classes. So 2/3 of all values cover two classes at the most. The accuracy of the calculation is in any case higher than the uncertainty of the characterization of the results in themselves.

10.2. Determination of the dominants



Picture 6 Holsti's Intra-Rater - Dominant lesion

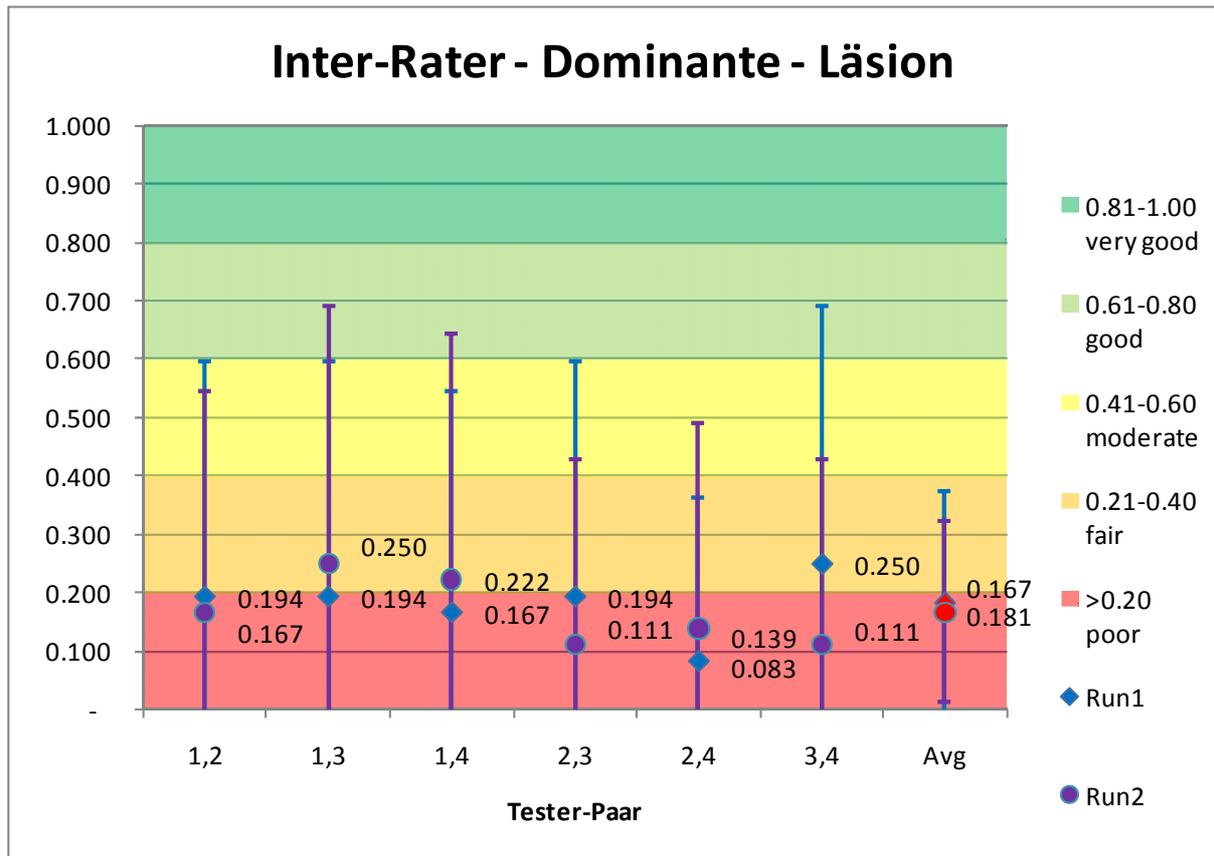
Avg = arithmetic average of the results of all coders

ICC = Reliability-coefficient according to Holsti (arithmetic average of the individual results of both runs of one coder)

The average results of the Intra-Rater-reliability for the determination of the dominant lesion of all four testers are located in the area moderate. For two testers the average is in the area $r_H : 0.6$, for two testers in the area $r_H : 0.3$.

The average Intra-Rater-reliability is with $r_H : 0.47 (\pm 0,501)$ also in the areamoderate. The wide standard deviation however shows that 2/3 of the values of the entirety are probably distributed in all classes and are therefore not evidential (Hüsler/Zimmermann, 2006).

The hypothesis 2.1. is considered as not confirmed. There is a moderate Intra-Rater-reliability for the determination of the dominant lesion of the spine by the LMO method. But this is not evidential because the dispersion is too big.



Picture 7 Holsti's Inter-Rater - Dominant lesion

Avg = arithmetic average of the results of all six coder-pairs (for both runs)

ICC = Reliability-coefficient according to Holsti (arithmetic average of the individual results of two testers in the same run. Two results for every tester-pair)

The average results of the Inter-Rater-reliability for the determination of the dominant lesion for all four testers are located in the area useless.

The average Inter-Rater-reliability is with r_H : 0.18 ($\pm 0,191$) also in the area useless. The great standard deviation shows that 2/3 of all values of the entirety are probably in the two lowest classes, and therefore are not evidential (Hüsler/Zimmermann, 2006).

The hypothesis 2.2 is to be considered as not confirmed. There is no Intra-Rater-reliability for the determination of the dominant lesion of the spine by the LMO method.