

**Interrater Reliabilitätstestung  
des Flexionstests der ersten Rippe bei  
symptomatischen und asymptomatischen  
Erwachsenen**

**MASTER - THESIS**

zur Erlangung des akademischen Grades

**Master of Science**

im Universitätslehrgang Osteopathie MSc

vorgelegt von

**Verena Kondler, BSc**

Matrikelnummer: 11946451

**Department für Gesundheitswissenschaften, Medizin und Forschung**

an der Donau-Universität Krems

Betreuerin 1: Mag. Dr. Astrid Grant Hay

Betreuer 2: Raimund Engel, MSc



29.09.2021

## EIDESSTATTLICHE ERKLÄRUNG

Ich, Verena Kondler, BSc, geboren am 16.12.1994 in Kirchdorf an der Krems, erkläre,

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

## DANKSAGUNGEN

Ein großes Dankeschön gilt meiner Masterthesis-Betreuerin Mag. Dr. Astrid Grant Hay, die mich mit ihrer hohen fachlichen Kompetenz und ihrer Begeisterung am wissenschaftlichen Arbeiten stets bestens betreut hat.

Danke auch an Christina Halasz, MSc, die als Studienleiterin fungiert und die Praxisräumlichkeiten zur Verfügung gestellt hat. Danke für die tatkräftige Unterstützung an der Umsetzung der Studie.

Darüber hinaus gebührt allen Probandinnen und Probanden ein großer Dank, welche durch ihr Mitwirken die Studie erst ermöglicht haben.

Meiner Schwester Magdalena Kondler, MA, meiner Freundin Elisabeth Schwab, MSc sowie meiner Freundin Magdalena Thallinger, BEd möchte ich für das Korrekturlesen des Konzeptes und der fertigen Arbeit ein großes Danke aussprechen.

Zu guter Letzt möchte ich meinem Partner, Arbeits- sowie Studienkollegen Maximilian Puhr, BSc für die gute Zusammenarbeit bei der Durchführung der Studie und das gegenseitige Motivieren zur Finalisierung der Thesis danken.

## ABSTRACT

### **Interrater Reliabilitätstestung des Flexionstests der ersten Rippe bei symptomatischen und asymptomatischen Erwachsenen**

**Ziel:** Der Flexionstest stellt einen häufig angewandten manualtherapeutischen Test zur Untersuchung der Mobilität der ersten Rippe dar. In der bisherigen Forschung wurde das Testverfahren noch nicht auf Reliabilität und Validität hin untersucht. Ziel dieser Arbeit ist es, Werte für die Interrater Reliabilität des Flexionstests zu erheben, um herauszufinden, wie hoch die Übereinstimmung zweier unabhängiger Testerinnen bzw. Tester ist.

**Methode:** Bei vorliegender Studie handelt es sich um eine Interrater Reliabilitätsstudie. Hierbei wurde der Flexionstest an 40 symptomatischen und asymptomatischen Erwachsenen von zwei unabhängigen Testerinnen bzw. Testern durchgeführt.

**Ergebnisse:** Aus den erhobenen Daten wurden mehrere Cohens Kappa Koeffizienten berechnet. Der Wert für alle Studienteilnehmerinnen und -teilnehmer liegt bei  $\kappa = 0.04$ , für Probandinnen bei  $\kappa = 0.03$  und für Probanden bei  $\kappa = 0.06$ . Alle Werte deuten auf eine geringfügige Übereinstimmung der Testerin bzw. des Testers hin.

**Konklusion:** Aufgrund der niedrigen Interrater Reliabilität des Flexionstests kann keine Empfehlung für dessen Verwendung in der klinischen Praxis abgegeben werden. Um eine finale Beurteilung des Tests abgeben zu können, wäre es essentiell auch dessen Intrarater Reliabilität sowie dessen Validität zu überprüfen.

**Schlüsselwörter:** Interrater Reliabilität, Flexionstest, erste Rippe, Osteopathie

## ABSTRACT

### **Interrater reliability of the flexion test for the first rib in symptomatic and asymptomatic adults**

**Objective:** The flexion test is a common examination test for the mobility of the first rib. Previous research has not yet investigated the reliability or validity of that test procedure. The aim of the present study is to collect data about the interrater reliability of the flexion test, in order to find out how two independent testers agree.

**Methods:** The present study is an interrater reliability design. Two independent testers examined 40 symptomatic and asymptomatic adults with the flexion test.

**Results:** The data was calculated with the Cohen's Kappa coefficient. For all participants the coefficient is  $\kappa = 0.04$ . Just for female participants it is  $\kappa = 0.03$  and for male participants it is  $\kappa = 0.06$ . All the values indicate a slight agreement between the testers.

**Conclusion:** The low interrater reliability does not allow a recommendation for the use of the flexion test in clinical practice. For a final appraisal of the test it would be relevant to examine the quality criterion of intrarater reliability and validity as well.

**Keywords:** interrater reliability, flexion test, first rib, osteopathy

# INHALTSVERZEICHNIS

<b>1 Einleitung</b> .....	<b>4</b>
<b>2 Theorie</b> .....	<b>6</b>
2.1 Gütekriterien von Messinstrumenten .....	6
2.2 Methodologische Grundlagen von Interrater Reliabilitätsstudien .....	7
2.2.1 Anzahl der Testerinnen bzw. Tester .....	7
2.2.2 Anzahl der Probandinnen und Probanden .....	7
2.2.3 Empfehlung der Verblindung .....	7
2.2.4 Empfehlung einer Trainingsphase .....	7
2.2.5 Statistische Berechnung der Ergebnisse .....	8
2.2.6 Interne und externe Validität einer Studie .....	8
2.3 Dysfunktionen der ersten Rippe .....	9
2.3.1 Definition .....	9
2.3.2 Symptome .....	9
2.3.2.1 Knöcherner Dysfunktionen .....	10
2.3.2.2 Respiratorische Dysfunktionen .....	10
2.3.2.3 Neurovaskuläre Dysfunktionen .....	10
2.3.3 Ursachen .....	11
2.3.4 Prävalenz .....	11
2.4 Mobilitätstestungen der ersten Rippe .....	12
2.4.1 Passive Methoden .....	12
2.4.2 Aktive Methoden .....	12
<b>3 Aktueller Forschungsstand</b> .....	<b>14</b>
3.1 Reliabilität von Mobilitätstests der ersten Rippe .....	14
3.2 Validität von Mobilitätstests der ersten Rippe .....	15
<b>4 Forschungsfrage und Hypothesen</b> .....	<b>16</b>
4.1 Forschungsfrage .....	16
4.2 Hypothesen .....	16
4.2.1 Nullhypothese $H_0$ .....	16
4.2.2 Alternativhypothese $H_1$ .....	16

<b>5 Methodik</b> .....	<b>17</b>
5.1 <i>Literaturrecherche</i> .....	17
5.2 <i>Studiendesign</i> .....	17
5.3 <i>Stichprobenbeschreibung</i> .....	18
5.3.1 <i>Einschlusskriterien</i> .....	18
5.3.2 <i>Ausschlusskriterien</i> .....	19
5.3.3 <i>Stichprobengröße</i> .....	20
5.3.4 <i>Rekrutierung der Stichprobe</i> .....	20
5.3.5 <i>Demographische Daten der Stichprobe</i> .....	20
5.3.6 <i>Daten zur Händigkeit</i> .....	20
5.3.7 <i>Daten zu vorliegender Symptomatik</i> .....	20
5.4 <i>Testerin und Tester</i> .....	21
5.5 <i>Studienleitung</i> .....	21
5.6 <i>Trainingsphase</i> .....	22
5.7 <i>Ablauf der Testung</i> .....	22
5.8 <i>Verblindung</i> .....	24
5.9 <i>Randomisierung</i> .....	24
5.10 <i>Markierung des Processus spinosus von Th1</i> .....	24
5.11 <i>Durchführung des Flexionstests</i> .....	26
5.12 <i>Statistische Datenaufbereitung</i> .....	27
5.12.1 <i>Darstellung der Daten</i> .....	27
5.12.2 <i>Berechnung und Interpretation des Cohens Kappa Koeffizienten</i> .....	28
5.12.3 <i>Berechnung der Häufigkeitsverteilungen der Testergebnisse</i> .....	29
5.12.4 <i>Beschreibung des Zusammenhangs zwischen Ergebnissen und Händigkeit</i> .....	29
<b>6 Ergebnisse</b> .....	<b>30</b>
6.1 <i>Kappa Koeffizient mit drei Ergebniskategorien</i> .....	30
6.2 <i>Kappa Koeffizient mit zwei Ergebniskategorien</i> .....	30
6.3 <i>Kappa Koeffizient mit drei Ergebniskategorien für Probandinnen</i> .....	31
6.4 <i>Kappa Koeffizient mit drei Ergebniskategorien für Probanden</i> .....	31
6.5 <i>Übersicht über alle errechneten Kappa Koeffizienten</i> .....	32
6.6 <i>Häufigkeitsverteilung der Testergebnisse</i> .....	32

6.7 Zusammenhang zwischen Ergebnissen und Händigkeit.....	34
6.8 Überprüfung der Hypothesen .....	34
<b>7 Diskussion .....</b>	<b>35</b>
7.1 Vergleich der Ergebnisse mit bestehender Literatur .....	35
7.2 Interpretation der Ergebnisse .....	35
7.2.1 Interpretation des Kappa Koeffizienten mit drei Ergebniskategorien .....	35
7.2.2 Interpretation des Kappa Koeffizienten mit zwei Ergebniskategorien .....	38
7.2.3 Interpretation des Kappa Koeffizienten mit drei Ergebniskategorien – geschlechterspezifisch .....	38
7.2.4 Interpretation der Häufigkeitsverteilung gefundener Testergebnisse .....	39
7.2.5 Interpretation des Zusammenhangs zwischen Ergebnissen und Händigkeit .....	39
7.3 Bedeutung der Ergebnisse im Kontext der Osteopathie .....	40
7.4 Kritische Reflexion der angewandten Methodik .....	40
7.4.1 Kritische Reflexion bezüglich Literaturrecherche .....	40
7.4.2 Kritische Reflexion bezüglich Studiendesign .....	41
7.4.3 Kritische Reflexion bezüglich Stichprobe .....	42
7.4.4 Kritische Reflexion bezüglich Studienpersonal .....	43
<b>8 Konklusion.....</b>	<b>44</b>
Literaturverzeichnis .....	46
Tabellenverzeichnis .....	50
Abbildungsverzeichnis .....	51
Abkürzungsverzeichnis .....	52
Anhang A – Studieninformation .....	53
Anhang B – Fragebogen bezüglich Ausschlusskriterien und Einverständniserklärung .....	55
Anhang C – Fragebogen bezüglich Händigkeit und Symptomatik .....	56
Anhang D – Test-Instruktion .....	57
Anhang E – Beurteilungsbogen .....	58
Anhang F – Rohdaten der Testung am 24.04.2021 .....	59
Anhang G – Englische Kurzfassung .....	60

# 1 Einleitung

Unspezifische Schmerzen im Bereich der Schulter können viele Ursachen haben. Sind die Beschwerden jedoch atemabhängig und gibt es keinerlei radiologische Auffälligkeiten, liegt die Vermutung nahe, dass es sich um eine Dysfunktion der ersten Rippe handeln könnte (Krumholz, 1998). Liegen zudem Traumata, insbesondere Schleudertraumata, oder eine Belastung durch das Tragen und Heben von einseitigen Lasten in der Anamnese vor, könnte sich der Verdacht einer Dysfunktion der ersten Rippe verhärtet (Mastromarchi & May, 2020). Eine Dysfunktion im Sinne einer Inspirationsstellung der ersten Rippe kann den Plexus brachialis mit seinen Nerven und Blutgefäßen komprimieren (Corts, 2019, S. 32). Daraus können ausstrahlende Schmerzen sowie Parästhesien im Bereich der oberen Extremität resultieren (Breidenbach & Erlewein, 2015, S. 72).

Somit wird eine Vielzahl an unterschiedlichen Symptomen der ersten Rippe zugeschrieben. Dies erfordert eine genaue und vor allem verlässliche Untersuchung der ersten Rippe, um eine mögliche Dysfunktion ausfindig machen zu können. Die manuelle Medizin bietet ein breites Spektrum an Untersuchungsmethoden für die erste Rippe an. Zugunsten eines verbesserten Überblicks können die Testverfahren in aktive und passive Herangehensweisen gegliedert werden.

In der aktuellen Literatur finden sich nur wenige Studien, welche die Zuverlässigkeit von passiven Mobilitätstests der ersten Rippe untersucht haben. Vier dieser Studien liefern jeweils Ergebnisse niedriger Interrater Reliabilität (Nomden et al., 2009; Smedmark et al., 2000; Fjellner et al., 1999; Heiderscheit & Boissonnault, 2008). Lindgren et al. (1992) hingegen fanden bei dem Cervical Rotation Lateral Flexion Test (= CRLF-Test) für die Beurteilung der Mobilität der ersten Rippe mit  $\kappa = 1.0$  eine perfekte Übereinstimmung zwischen den Untersuchungspersonen. Einen methodologisch limitierenden Faktor dieser Studie stellt jedoch die geringe Anzahl der Probandinnen und Probanden dar.

Aktive Mobilitätstests, wie der Flexionstest, wurden laut aktuellem Forschungsstand bisher nicht auf ihre Verlässlichkeit hin überprüft. Der Flexionstest der ersten Rippe zeichnet sich, verglichen mit dem CRLF-Test, durch eine simple und praktikable Anwendbarkeit aus. Aufgrund der Durchführung in sitzender Ausgangsposition ist der Test gut in den Untersuchungsablauf in der osteopathischen Praxis integrierbar. Zudem ermöglicht die bilaterale Durchführung des Flexionstests eine Aussage über beide ersten Rippen im direkten Seitenvergleich. Der CRLF-Test hingegen wird unilateral angewandt, was einen erhöhten Zeitaufwand mit sich bringt, sofern beide erste Rippen im Seitenvergleich beurteilt werden sollen.

Peeters und Lason (2015, S.79) beschreiben für die Durchführung des Flexionstests keine Kontraindikationen, welche gegen ein Applizieren des Tests sprechen würden. Folglich kann der Mobilitätstest bei einem Großteil der Patientinnen und Patienten im klinischen Alltag angewandt werden.

Die bis dato fehlende Untersuchung hinsichtlich der Gütekriterien Reliabilität und Validität des Flexionstests der ersten Rippe stellt eine Forschungslücke dar und soll als Anlass für diese Masterthesis genommen werden. Das Ziel der vorliegenden Arbeit ist es, die Interrater Reliabilität des Flexionstests der ersten Rippe zu untersuchen.

## 2 Theorie

Dieses Kapitel soll Aufschluss über die theoretischen Grundlagen des Forschungsgegenstandes sowie den methodologischen Aufbau von Interrater Reliabilitätsstudien geben.

### 2.1 Gütekriterien von Messinstrumenten

Messinstrumente unterliegen mehreren zu erfüllenden Gütekriterien. Hussy et al. (2013, S. 23) fassen die in der Literatur bestehenden Kriterien auf folgende drei zusammen: Objektivität, Reliabilität und Validität.

Objektivität ist mit der Unabhängigkeit der Testergebnisse von externen Einflüssen, wie beispielsweise der Testumgebung sowie den Untersuchungspersonen, gleichzusetzen. Dies wird über eine möglichst detaillierte Standardisierung von Durchführung, Auswertung und Interpretation des Testverfahrens erreicht. Das Kriterium der Objektivität stellt eine Voraussetzung für die Erfüllung der weiteren Gütekriterien dar (Hussy et al., 2013, S. 23f).

Reliabilität wird definiert als Verlässlichkeit eines Messinstruments und beschreibt demnach, ob mehrmalige Messungen an derselben Stichprobe gleiche Ergebnisse erzielen (Scherfer et al., 2011, S. 244). Dabei werden zwei unterschiedliche Arten beschrieben: Interrater Reliabilität und Intrarater Reliabilität.

Die Interrater Reliabilität beschreibt die Übereinstimmung zweier oder mehrerer Untersuchender bei der Durchführung eines Testverfahrens. Bei der Intrarater Reliabilität hingegen wird die Übereinstimmung bei wiederholter Testdurchführung durch ein und denselben Untersuchenden thematisiert (Scherfer et al., 2011, S. 250).

Jene Messinstrumente, welche nicht den Anforderungen der Reliabilität entsprechen, können folglich auch nicht valide sein (Scherfer et al., 2011, S. 244).

Validität wird mit dem Begriff Gültigkeit gleichgesetzt und beschreibt, ob ein Messinstrument tatsächlich das misst, was es zu messen vorgibt (Scherfer et al., 2011, S. 242f).

Scherfer et al. (2011, S. 243) ergänzen die Liste der zu erfüllenden Gütekriterien um ein weiteres Kriterium – jenes der Praktikabilität. Hierbei stellt sich die Frage, wie gut das Messinstrument im klinischen Alltag angewandt werden kann. Es werden sowohl ethische als auch ökonomische Aspekte beleuchtet. Neben dem zeitlichen Aufwand und der möglichen gesundheitlichen Belastung der Patientin bzw. des Patienten darf auch der durch den Test entstehende zeitliche Aufwand für die Therapeutin bzw. den Therapeuten nicht außer Acht gelassen werden (Scherfer et al., 2011, S. 246).

## **2.2 Methodologische Grundlagen von Interrater Reliabilitätsstudien**

Interrater Reliabilitätsstudien zählen zur übergeordneten Gruppe der diagnostischen Studien, welche Messinstrumente auf ihre Gütekriterien hin überprüfen (Scherfer et al., 2011, S. 247). Die International Academy of Manual/Musculoskeletal Medicine hat Empfehlungen für Reliabilitäts- und Validitätsstudien herausgebracht (Patijn, 2019), welche in diesem Kapitel genauer beschrieben werden sollen. Zudem werden die Begrifflichkeiten der internen und externen Validität in Bezug auf die Qualität von Studien definiert.

### **2.2.1 Anzahl der Testerinnen bzw. Tester**

Laut Patijn (2019) reichen zwei Testerinnen bzw. Tester für die Durchführung einer Reliabilitätsstudie aus. Durch jede weitere Person könnten aufgrund des oftmaligen Durchführens des Tests therapeutische Effekte hervorgerufen werden, welche infolgedessen die Aussagekraft des Ergebnisses stark beeinflussen würden (Patijn, 2004).

Die Notwendigkeit von mehreren Untersucherinnen bzw. Untersuchern für Reliabilitätsstudien stellt sich ausschließlich bei einer gewünschten Evaluierung des Ausbildungsstandes der einzelnen Personen bzw. bei der Evaluierung des Effekts von vorangegangenen Trainingsphasen (Patijn, 2019).

### **2.2.2 Anzahl der Probandinnen und Probanden**

Die International Academy of Manual/Musculoskeletal Medicine betrachtet eine Anzahl von 40 Probandinnen und Probanden für eine Reliabilitätsstudie als ausreichend (Patijn, 2019).

### **2.2.3 Empfehlung der Verblindung**

Eine Verblindung zwischen den einzelnen Testerinnen bzw. Testern ist für die Qualität einer Reliabilitätsstudie ebenso essenziell wie eine Verblindung zwischen den Testerinnen bzw. Testern und den Getesteten (Patijn, 2019). Dafür empfiehlt der Autor, die Kommunikation zwischen den beiden Testerinnen und Testern sowie zwischen den Testerinnen bzw. Testern und den Getesteten auf ein Minimum zu reduzieren. Zudem sollten die Testerinnen bzw. Tester eigene Beurteilungsbögen verwenden, um so eine mögliche Beeinflussung durch das Ergebnis der zweiten Person zu vermeiden. Diese Beurteilungsbögen sollten im Anschluss an die Testung von einer unabhängigen Studienleitung eingesammelt werden.

### **2.2.4 Empfehlung einer Trainingsphase**

Patijn (2019) setzt der tatsächlichen Studie eine Phase der Standardisierung voraus. In dieser Phase soll eine Einigung des Studienpersonals bezüglich exakter standardisierter Durchführung und Beurteilung des Testverfahrens erfolgen. Diese Trainingsphase wird für

jede Testerin bzw. jeden Tester, ungeachtet des aktuellen Ausbildungsstandes und der vorhandenen Berufserfahrung, empfohlen und kann in zwei Phasen unterteilt werden:

- gemeinsame Trainingsphase: Diese wird von beiden Untersuchenden gemeinsam durchlaufen und soll helfen, den geplanten Ablauf zu kontrollieren.
- klinischer Alltag: Den Untersuchenden wird empfohlen, das standardisierte Testverfahren in den Arbeitsalltag zu integrieren, um die eigenen Fähigkeiten zu schulen und um möglichst sicher in der Ausführung des Testverfahrens zu werden.

Wird die Durchführung dieser Trainingsphase unterlassen, repräsentiert der errechnete Cohens Kappa Koeffizient laut Patijn (2019) schlussendlich überwiegend die individuelle Interpretation des Testverfahrens durch die Testerinnen bzw. Tester und nicht wie gewünscht die Verlässlichkeit des Testverfahrens.

### **2.2.5 Statistische Berechnung der Ergebnisse**

Zur statistischen Berechnung des Zusammenhangs der Ergebnisse von Reliabilitätsstudien gibt es je nach Messniveau unterschiedliche Koeffizienten. Während für intervallskalierte und ratioskalierte Messniveaus neben dem Intraklassen-Korrelations-Koeffizienten auch der Pearson'sche Korrelationskoeffizient  $r$  herangezogen werden kann, soll bei einem ordinalen Messniveau der Rangkorrelationskoeffizient nach Spearman berechnet werden (Scherfer et al., 2011, S. 252). Auf nominalem Messniveau wird der Cohens Kappa Koeffizient zur Berechnung der Reliabilität empfohlen (Patijn, 2019). Mithilfe dieses Wertes ist es möglich eine Aussage darüber zu treffen, wie oft die Testerinnen bzw. Tester übereinstimmen und vor allem wie weit diese Übereinstimmung über den Zufall hinausgeht (Scherfer et al., 2011, S. 252).

Neben dem Cohens Kappa Koeffizienten steht für nominale Daten noch eine weitere Berechnungsmöglichkeit zur Verfügung. Die Variable  $p_o$  beschreibt die prozentuelle Übereinstimmung der Testerinnen bzw. Tester bei allen Ergebnissen. Als nachteilig gegenüber dem Cohens Kappa Koeffizienten erweist sich hierbei die Tatsache, dass durch den Wert die Übereinstimmung der Untersuchenden überschätzt wird (Döring & Bortz, 2016, S.567), weil die durch Zufall entstandenen Übereinstimmungen nicht in die Berechnung des Wertes miteinbezogen werden (Patijn, 2019).

### **2.2.6 Interne und externe Validität einer Studie**

Ein hohes Maß an Validität einer Studie liegt dann vor, wenn die Ergebnisse nicht durch Verzerrungen beeinflusst worden sind (Scherfer et al., 2011, S. 154). Dabei können zwei Arten von Validität unterschieden werden – die interne und die externe Validität (Schäfer & Schöttker-Königer, 2015, S. 5).

Interne Validität beschreibt, inwiefern externe Störfaktoren einen Einfluss auf das Studienergebnis haben. Um einen möglichst hohen Grad an interner Validität erreichen zu können, sollen diese externen Faktoren so gering wie nur möglich gehalten werden bzw. bedarf es ihrer Kontrolle (Schäfer & Schöttker-Königer, 2015, S. 5). Hierfür empfiehlt es sich, die Rahmenbedingungen für die Durchführung der Studie weitgehend zu standardisieren (Schäfer & Schöttker-Königer, 2015, S. 5), zum Beispiel über die Randomisierung der Studienteilnehmenden sowie durch die Verblindung von Studienteilnehmenden und -personal (Scherfer et al., 2011, S.162).

Externe Validität kann mit dem Begriff der Übertragbarkeit erklärt werden und wird dann als hoch angesehen, wenn die Ergebnisse gut auf den klinischen Alltag übertragen werden können (Schäfer & Schöttker-Königer, 2015, S. 5). Dies kann von Seiten der Studienautorin bzw. des -autors unter anderem über die Wahl des Settings und auch über die Wahl der Ein- und Ausschlusskriterien beeinflusst werden (Scherfer et al., 2011, S. 157).

Eine hohe interne Validität geht mit einer niedrigen externen Validität einher und vice versa (Schäfer & Schöttker-Königer, 2015, S. 6). Folglich steht die Anforderung an eine Studie, einen möglichst hohen Grad an interner Validität zu erreichen, dem Anspruch, einen möglichst hohen Grad der externen Validität zu erreichen, gegenüber.

## **2.3 Dysfunktionen der ersten Rippe**

In diesem Kapitel soll das klinische Bild von Dysfunktionen der ersten Rippe detailliert dargestellt werden.

### **2.3.1 Definition**

Michael (1995) definiert eine Dysfunktion der ersten Rippe als beeinträchtigte Beweglichkeit. Die Rippe kann demnach in einer Elevationsstellung (= Inspirationsdysfunktion) oder in einer Depressionsstellung (= Expirationsdysfunktion) fixiert sein. Im Bereich der ersten Rippe treten laut Autor Inspirationsdysfunktionen häufiger auf.

### **2.3.2 Symptome**

Die aufgrund einer Dysfunktion der ersten Rippe hervorgerufenen Symptome äußern sich häufig in Form von Schmerzen im posterioren subscapulären oder anterioren subclaviculären Raum (Krumholz, 1998). Der Autor beschreibt weiter, dass sich die Schmerzen auf der linken Seite wie Beschwerden bei Angina pectoris präsentieren können. Zudem können Dysfunktionen der ersten Rippe häufig Verspannungen im Musculus trapezius hervorrufen (Reichert, 2018, S. 343).

Die aufgrund einer Funktionsstörung der ersten Rippe entstehenden Probleme sind laut Michael (1995) vielfältig und können demzufolge zur besseren Übersicht in drei Hauptbereiche gegliedert werden: knöcherne, respiratorische und neurovaskuläre Dysfunktionen.

### **2.3.2.1 Knöcherne Dysfunktionen**

Auffälligkeiten der ersten Rippen können Dysfunktionen im Bereich des ersten Brustwirbels (= Th1), des Manubrium sterni, der Halswirbelsäule und des Schultergelenks verursachen (Michael, 1995), welche jedoch nicht zwangsläufig im Röntgen sichtbar sein müssen (Krumholz, 1998).

Die Symptome einer Blockade der ersten Rippe können jenen Symptomen, welche durch ein blockiertes Segment des cervikothorakalen Übergangs (Segmente C6, C7 und Th1) hervorgerufen werden, ähneln (Breidenbach & Erlewein, 2015, S. 72).

### **2.3.2.2 Respiratorische Dysfunktionen**

Beschwerden im Zusammenhang mit der ersten Rippe können sich atemabhängig präsentieren – folglich können sich die Symptome je nach Art der Dysfunktion bei Ein- oder Ausatmung verstärken (Krumholz, 1998). Anstrengende Tätigkeiten wie das Heben oder Tragen von Lasten können ebenfalls zu einer Aggravation der Symptome führen, in Ruhe setzt zumeist Linderung ein (Michael, 1995).

Eine Erklärung für die respiratorischen Auswirkungen einer Dysfunktion der ersten Rippe könnte das Ligamentum costopleurale darstellen, über welches eine enge Verbindung zwischen der Pleura und der ersten Rippe besteht (Corts, 2019, S. 26).

### **2.3.2.3 Neurovaskuläre Dysfunktionen**

Eine Dysfunktion im Sinne einer Inspirationsstellung der ersten Rippe kann den Plexus brachialis mit seinen Nerven und Blutgefäßen sowohl komprimieren (Corts, 2019, S. 32), als auch unter Zug bringen (Hooper et al., 2010). Auch Laekeman und Kreutzer (2009, S. 293) zeigen eine Inspirationsstellung als mögliche Ursache für das Auftreten eines Thoracic Outlet Syndromes (= TOS) auf. Dieses Kompressionsphänomen der ersten Rippe kann sich je nach komprimierter Struktur in Form von ausstrahlenden Schmerzen oder Parästhesien im Bereich der oberen Extremität auswirken (Breidenbach & Erlewein, 2015, S. 72). Einer möglichen Kompression durch die erste Rippe sind folgende Strukturen ausgesetzt: Plexus brachialis, Vena und Arteria subclavia (Corts, 2019, S. 32) sowie Ductus thoracicus, Ductus lymphaticus und Venae jugulares, was folglich eine Beeinträchtigung des venolymphatischen Abflusses mit sich bringen kann (Mitchell & Mitchell, 2004, S. 86).

Je nachdem, welche Struktur durch eine Dysfunktion der ersten Rippe betroffen ist, können zusätzlich folgende Symptome in der oberen Extremität auftreten (Hooper et al., 2010):

- Eine arterielle Kompression kann Blässe und ein Gefühl von Kälte hervorrufen.
- Zyanose, Schweregefühl und Ödeme stellen potenzielle Hinweise auf eine venöse Kompression dar.
- Eine Kompression der nervalen Strukturen kann in Taubheit, Schwäche sowie Reduktion der Feinmotorik resultieren.

Zudem ist eine Mitbeteiligung des sympathischen Grenzstranges (Michael, 1995) und des Ganglion stellatum mit dadurch hervorgerufenen autonomen Reaktionen im Bereich von Gesicht und oberer Extremität möglich (Corts, 2019, S. 32).

### **2.3.3 Ursachen**

Als mögliche Auslöser für Dysfunktionen der ersten Rippe werden in der Literatur Traumata, insbesondere Schleudertraumata, sowie das einseitige Tragen oder Heben von schweren Lasten beschrieben (Mastromarchi & May, 2020). Eine Blockade auf Höhe des ersten Brustwirbels kann ebenfalls in einer Dysfunktion der ersten Rippe resultieren (Hüter-Becker & Dölken, 2015, S. 224). Michael (1995) beschreibt weiter Haltungsasymmetrien im Bereich des cervikothorakalen Übergangs und Schultergürtels als mögliche Ursache für Dysfunktionen der ersten Rippe.

Corts (2019, S. 26) macht zudem auf den Zusammenhang zwischen Lunge und erster Rippe aufmerksam. Hierfür nennt die Autorin eine überstandene Pneumonie als Beispiel, welche aufgrund reduzierter Gewebeelastizität über die Membrana suprapleuralis einen Zug auf die erste Rippe erzeugen kann und damit ursächlich für eine Dysfunktion sein könnte.

### **2.3.4 Prävalenz**

Bezüglich der Prävalenz von Dysfunktionen der ersten Rippe finden sich in der Literatur keine Publikationen.

In der qualitativen Studie von Mastromarchi und May (2020) wurden jedoch Expertinnen und Experten zur Häufigkeit des Auftretens von Dysfunktionen der ersten Rippe in der klinischen Praxis befragt. Laut ihnen hat die erste Rippe bei 20 bis 30% der Patientinnen und Patienten mit Schulter- oder Nackenproblemen einen Einfluss auf vaskuläre und neurodynamische Faktoren dieser Beschwerden.

## **2.4 Mobilitätstestungen der ersten Rippe**

In der manuellen Medizin stehen, wie eingangs erwähnt, viele verschiedene Untersuchungsmethoden für die Mobilität der ersten Rippe in diversen Ausgangsstellungen zur Verfügung. Laut Patijn (2019) reicht in der Praxis selten ein einzelner Test für eine sichere Diagnosestellung einer Dysfunktion aus. Demzufolge empfiehlt es sich auch im Falle der ersten Rippe, zur Mobilitätstestung nicht ausschließlich den Flexionstest, sondern mehrere Mobilitätsuntersuchungen anzuwenden.

Die Methoden können zur besseren Übersicht in passive und aktive Tests eingeteilt werden.

### **2.4.1 Passive Methoden**

Die passive Mobilität der ersten Rippe wird häufig mittels Springing-Test oder Federtest in Rückenlage (Liem & Dobler, 2002, S. 272) bzw. im Sitz (Reichert, 2018, S. 343) getestet. Hierbei kann durch einen nach caudal gerichteten Druck auf die erste Rippe die Mobilität beurteilt werden.

Zudem wird in der Literatur auch das Applizieren von posterior-anterior gerichtetem Druck (PA) auf die Rippe in Bauchlage und anterior-posterior gerichtetem Druck (AP) in Rückenlage beschrieben (Heiderscheit & Boissonnault, 2008).

Der CRLF-Test untersucht eine mögliche Elevationsstellung der ersten Rippe, bei dem die Testerin bzw. der Tester den Kopf der Probandin bzw. des Probanden passiv in eine maximale kontralaterale Rotation und ipsilaterale Lateralflexion bewegt und dabei das Endgefühl bewertet (Lindgren et al., 1992).

Strunk (2012, S. 102) und Corts (2016, S. 101) beschreiben Mobilitätstests, die nicht ausschließlich die erste Rippe beurteilen, sondern Auskunft über die Elastizität der gesamten oberen Thoraxapertur geben.

### **2.4.2 Aktive Methoden**

Die aktiven Tests umfassen das Einbeziehen der Atmung in Rückenlage (Liem & Dobler, 2002, S. 278) oder im Sitz (Langer & Hebgen, 2012, S. 78; Ward, 2003, S. 722; Hüter-Becker & Dölken, 2015, S. 223). Hierbei wird die Mobilität der ersten Rippe bei Inspiration und Expiration untersucht.

Eine Beurteilung der Mobilität der ersten Rippe ist auch durch das zusätzliche Bewegen des Kopfes möglich. Während bei Peeters und Lason (2015, S. 79) eine Flexion des Kopfes eingeleitet wird, beurteilt Krumholz (1998) die erste Rippe mittels einer Hyperextensions-Rotationsbewegung des Kopfes.

Eine Testvariante kombiniert das Einbeziehen der Atmung mit dem aktiven Bewegen des Kopfes in Flexion und Extension (Mitchell & Mitchell, 2004, S. 91f).

Eine zusätzliche Bewegung der Arme kann ebenfalls zur Beurteilung der Mobilität der ersten Rippe herangezogen werden. Dabei wird entweder ein Arm aktiv in Flexion bewegt (Streeck et al., 2006, S. 357) oder beide Arme zugleich (Krumholz, 1998).

Anschließend soll insbesondere der Flexionstest von Peeters und Lason (2015, S. 79) genauer beleuchtet werden. Die Autoren beschreiben den Flexionstest der ersten Rippe erstmals in der Literatur. Die Ausgangsstellung für die Testung ist der entspannte Sitz. Die Testerin bzw. der Tester steht hinter der Probandin bzw. dem Probanden und palpiert mit den Daumen beidseits den Angulus costae der ersten Rippen von posterior. Anschließend wird der Kopf aktiv flektiert und die Daumen folgen der Bewegung der ersten Rippen nach cranial. Bewegt eine Seite schneller als die andere, liegt eine Dysfunktion dieser Seite vor.

In der Testbeschreibung werden keinerlei Kontraindikationen für die Anwendung der Testung erwähnt.

### 3 Aktueller Forschungsstand

Im Zuge der systematischen Literaturrecherche wurde die Forschungslandschaft auf Daten bezüglich Reliabilität und Validität von Mobilitätstests der ersten Rippe hin durchsucht.

#### 3.1 Reliabilität von Mobilitätstests der ersten Rippe

Einige der passiven Mobilitätstests der ersten Rippe wurden bereits auf ihre Interrater Reliabilität hin untersucht.

Während Fjellner et al. (1999) die Mobilität der ersten Rippe im Sitz untersucht haben, wurde bei Smedmark et al. (2000) die Rückenlage als Ausgangsposition für die Testung der Mobilität mittels Springing gewählt. Nomden et al. (2009) wählten als Untersuchung der Mobilität der ersten Rippe ebenfalls die Springing-Methode, bezüglich Ausgangsstellung findet sich jedoch keine Beschreibung in der Studie. Heiderscheit und Boissonnault (2008) wählten als Untersuchung der Mobilität der ersten Rippe die Applikation von PAs sowie APs. Lindgren et al. (1992) untersuchten den CRLF-Test auf seine Interrater Reliabilität.

Vier dieser Studien liefern jeweils Ergebnisse niedriger Interrater Reliabilität (Nomden et al., 2009; Smedmark et al., 2000; Fjellner et al., 1999; Heiderscheit & Boissonnault, 2008).

Bei Smedmark et al. (2000) und Nomden et al. (2009) wurden Cohens Kappa Werte von  $\kappa = 0.35$  bzw.  $\kappa = 0.26$  gemessen. Landis und Koch (1977, S. 165) interpretieren Kappa Werte zwischen 0.21 und 0.40 als mäßige Übereinstimmung der Testergebnisse.

Fjellner et al. (1999) sowie Heiderscheit und Boissonnault (2008) werden demnach mit ihren Kappa Werten von  $\kappa = 0.06$  bzw.  $\kappa = 0.11$  lediglich der Kategorie der geringen Übereinstimmung zugeteilt.

Lindgren et al. (1992) konnten bei ihrer Studie einen Kappa Wert von  $\kappa = 1.0$  erreichen, was einer perfekten Übereinstimmung zweier Untersuchender gleichkommt. Einen methodologisch limitierenden Faktor dieser Studie stellt jedoch die geringe Anzahl der Studienteilnehmerinnen und -teilnehmer von 23 dar. Patijn (2019) empfiehlt für einen aussagekräftigen Outcome dieses Studiendesigns 40 Probandinnen und Probanden zu testen.

Zur besseren Übersicht werden die Ergebnisse der genannten Studien in untenstehender Tabelle 1 aufgelistet.

**Tabelle 1.** Kappa Koeffizienten für Mobilitätstests der ersten Rippe in der Literatur

Studien	Mobilitätstest	Kappa Koeffizient
Fjellner et al., 1999	Mobilitätstestung im Sitz	0.06
Heiderscheit & Boissonnault, 2008	Applikation von PAs sowie APs	0.11
Nomden et al., 2009	Springing	0.26
Smedmark et al., 2000	Springing in Rückenlage	0.35
Lindgren et al., 1992	CRLF-Test	1.0

Aktive Mobilitätstests, wie der Flexionstest, wurden laut aktuellem Forschungsstand bisher nicht auf ihre Verlässlichkeit untersucht. Die Recherche ergab dazu keinerlei Angaben, weder über die Intrarater Reliabilität noch über die Interrater Reliabilität des Flexionstests für die erste Rippe.

### **3.2 Validität von Mobilitätstests der ersten Rippe**

Die in Kapitel 2.4 beschriebenen Mobilitätstests für die erste Rippe wurden bis auf den CRLF-Test bisher noch nicht auf ihre Validität hin untersucht.

Vaught et al. (2011) errechneten basierend auf der Studie von Lindgren et al. (1992) für den CRLF-Test eine Sensitivität von 93% und eine Spezifität von 94%. Als Goldstandard diente Lindgren et al. (1992) das Prozedere der Röntgenkinematographie.

## **4 Forschungsfrage und Hypothesen**

Folgend sollen die zu beantwortende Forschungsfrage sowie die zu untersuchenden Hypothesen angeführt werden, auf denen vorliegende Arbeit aufbaut.

### **4.1 Forschungsfrage**

Die Forschungsfrage dieser Studie lautet wie folgt: Wie hoch ist die Übereinstimmung der Ergebnisse des Flexionstests bei symptomatischen sowie asymptomatischen Erwachsenen durchgeführt von zwei unabhängigen Untersuchungspersonen?

### **4.2 Hypothesen**

Im Zuge der Arbeit wurden eine Null- und eine Alternativhypothese aufgestellt.

#### **4.2.1 Nullhypothese $H_0$**

Die Übereinstimmung der Ergebnisse des Flexionstests für die erste Rippe ist nicht substanziell (Landis & Koch, 1977, S. 165), der Cohens Kappa Koeffizient  $\kappa$  ist somit kleiner als 0.61.

#### **4.2.2 Alternativhypothese $H_1$**

Die Übereinstimmung der Ergebnisse des Flexionstests für die erste Rippe ist mindestens substanziell (Landis & Koch, 1977, S. 165), der Cohens Kappa Koeffizient  $\kappa$  ist somit größer bzw. gleich 0.61.

## 5 Methodik

Orientierend an den in Kapitel 2.2 angeführten Anforderungen an eine Reliabilitätsstudie wird nachfolgend der methodische Aufbau der vorliegenden Studie beschrieben.

### 5.1 Literaturrecherche

Die systematische Literaturrecherche zum aktuellen Forschungsstand fand im Zeitraum von 16.11. bis 23.11.2020 statt. In folgenden medizinischen Datenbanken wurde gesucht: PubMed, PEDro, Cochrane Library, Osteopathic Research Web, Cambase und Ostmed.

Datenbanken für elektronisch zugängliche Journals wurden ebenfalls durchsucht:

- [chiromt.com](http://chiromt.com)
- [ptjournal.apta.org](http://ptjournal.apta.org)

Um die Verzerrung durch einen Retrieval Bias zu minimieren (Scherfer et al., 2011, S. 60), wurde mit möglichst vielen Begrifflichkeiten und deren Synonymen in Verbindung mit verschiedenen Booleschen Operatoren (AND und OR) gesucht:

- rib
- reliability / reproducibility of results / reproducibility of findings / interjudge reliability / interobserver reliability / intertester reliability / interexaminer reliability / interrater reliability
- motion / mobility / palpation / diagnostic / manual examination / physical examination / mobility test / motion test / mobility assessment / motion assessment / range of motion / articular

Bei der Datenbanksuche wurden die Filter human, german und english verwendet. Ergänzend zu der systematischen Literatursuche wurde anschließend eine heuristische Suchstrategie verfolgt.

### 5.2 Studiendesign

Bei vorliegender Studie handelt es sich um eine methodologische Studie, präziser um ein quantitatives Design in Form einer Interrater Reliabilitätsstudie. Im Zuge dieser Studie wurde an derselben Stichprobe neben der Interrater Reliabilität des Flexionstests auch die Interrater Reliabilität des costoclaviculären Manövers ermittelt. Hierbei handelt es sich um einen Test zur Untersuchung von TOS. Die Testung wird in der Masterthesis von Maximilian Pühr beschrieben.

## **5.3 Stichprobenbeschreibung**

Im folgenden Abschnitt werden neben den Ein- und Ausschlusskriterien auch die Größe der Stichprobe sowie die Rekrutierungswege detailliert beschrieben. Ebenso befindet sich eine demographische Analyse der Stichprobe im Kapitel.

### **5.3.1 Einschlusskriterien**

In der Literatur finden sich widersprüchliche Empfehlungen bezüglich der Frage, ob in Reliabilitätsstudien symptomatische und/oder asymptomatische Probandinnen und Probanden untersucht werden sollen.

Van Trijffel et al. (2005) stellen in ihrem systematischen Review über Reliabilitätsstudien fest, dass Studien, welche symptomatische Probandinnen und Probanden untersucht haben, häufig bessere Ergebnisse erzielen als Studien mit asymptomatischen Teilnehmenden. Die Autorinnen und Autoren empfehlen, die Untersuchung von repräsentativen und somit symptomatischen Probandinnen und Probanden, um dem Anspruch der externen Validität zu genügen.

Laut Patijn (2019) müssen für die Umsetzung einer Reliabilitätsstudie hingegen nicht zwangsläufig symptomatische Studienteilnehmerinnen und -teilnehmer mit Beschwerden in der getesteten Region herangezogen werden. Der Autor begründet dies mit der Tatsache, dass bei vorliegendem Studiendesign nicht die Existenz einer Diagnose bestätigt, sondern vielmehr die Reproduzierbarkeit eines Testverfahrens sowie dessen endgültige Interpretation durch die Testerin bzw. den Tester untersucht werden soll.

Fjellner et al. (1999) begründen ihre Entscheidung für die Wahl von ausschließlich asymptomatischen Probandinnen und Probanden mit der Tatsache, dass Patientinnen und Patienten mit Schmerzen häufig sehr empfindlich sind und dadurch manuelle Untersuchungen oftmals nicht durchführbar sind, was demzufolge die Aussagekraft einer Studie herabsetzen würde.

Heiderscheit und Boissonnault (2008) wählten in ihrer Reliabilitätsstudie über die Beweglichkeitstestung des Thorax ebenso bewusst asymptomatische Probandinnen und Probanden. Dies begründen sie folgendermaßen: Untersuchungen des Brustkorbes werden sowohl bei Symptomen im Brustkorb als auch bei Patientinnen und Patienten mit Beschwerden in angrenzenden Körperregionen, wie der Schulter oder der Halswirbelsäule, angewandt. Ein Symptom in der zu untersuchenden Körperregion ist folglich nicht zwingend notwendig für die Teilnahme an einer Reliabilitätsstudie. Diese Begründung lässt sich auch auf die Untersuchung der ersten Rippe umlegen. Der Flexionstest wird in der täglichen Praxis nicht ausschließlich bei symptomatischen (die erste Rippe betreffend) Patientinnen und Patienten durchgeführt. Vielmehr wird er auch bei symptomfreien Patientinnen und Patienten,

welche möglicherweise Beschwerden in angrenzenden Körperregionen, wie Schulter, Halswirbelsäule oder Viszerum aufweisen, angewandt. Auch bei Problemen die Atmung oder den venösen, arteriellen sowie lymphatischen Kreislauf betreffend, ist eine Untersuchung des knöchernen Thorax von hoher Relevanz (Greenman, 2005, S. 273).

Aufgrund oben angeführter Argumentation soll die gewählte Stichprobe sowohl symptomatische als auch asymptomatische Erwachsene miteinbeziehen. Somit wird eine möglichst praxisnahe Population repräsentiert, um eine hohe externe Validität der Studie und somit eine gute Übertragbarkeit der Ergebnisse in den klinischen Alltag gewährleisten zu können.

Das Mindestalter wurde auf 18 Jahre festgelegt. Bis zu diesem Alter sollten die Knochenreifung und somit das Wachstum beendet sein (Matzen et al., 2007, S. 138).

### **5.3.2 Ausschlusskriterien**

Die Ausschlusskriterien wurden mittels Fragebogen vor Studienantritt abgeklärt. Folgende Kriterien schlossen eine Teilnahme an der Studie aus:

- akute knöcherner Verletzungen (< 6 Wochen) in folgenden Bereichen: HWS, Clavicula, BWS, Rippen und gesamte obere Extremität, aufgrund der noch nicht abgeschlossenen Wundheilung (Diemer & Sutor, 2011, S. 56)
- ärztlich vorgegebene Bewegungseinschränkungen der Wirbelsäule bzw. der oberen Extremität, wodurch die Ausgangs- bzw. Endposition der Testungen nicht eingenommen werden soll
- Probandinnen und Probanden, welche von einer der beiden oder beiden Untersuchungspersonen in vorangegangenen Therapieeinheiten bereits untersucht und behandelt wurden, um die Verblindung der Studie nicht zu gefährden

Nachdem im Zuge der Studie neben dem Flexionstest für die erste Rippe auch das costoclaviculäre Manöver durchgeführt wurde, erweiterte sich die Liste der Ausschlusskriterien um folgenden Aspekt:

- In der Literatur gibt es bezüglich der Durchführung des costoclaviculären Manövers bei vorliegenden Zusatzdiagnosen, wie Aneurysma und Herzinsuffizienz, keinerlei Empfehlungen bzw. Warnungen. Aus Sicherheitsgründen wurden in dieser Studie jedoch Probandinnen und Probanden mit einem diagnostizierten Aneurysma oder Herzinsuffizienz ausgeschlossen, um durch die Testungen keinen zusätzlichen Druck auf das Herzkreislaufsystem auszulösen.

### **5.3.3 Stichprobengröße**

Wie in Kapitel 2.2.2 beschrieben, wurde für die Durchführung vorliegender Reliabilitätsstudie eine Anzahl von 40 Probandinnen und Probanden angestrebt und auch erreicht. Bei keiner bzw. keinem der 40 Probandinnen und Probanden musste die Testung vorzeitig abgebrochen werden.

### **5.3.4 Rekrutierung der Stichprobe**

Die Rekrutierung der Teilnehmerinnen und Teilnehmer erfolgte einerseits über eine Ausschreibung für die Mitarbeiterinnen und Mitarbeiter des Rehabilitationszentrums Gröbming der PVA. Andererseits wurden Probandinnen und Probanden aus dem Bekannten- und Verwandtenkreis des Studienpersonals rekrutiert.

Interessierte meldeten sich per Mail bei der Studienleitung. Der Aufklärungsbogen und der Fragebogen zur Abklärung der Ausschlusskriterien wurde den Teilnehmerinnen und Teilnehmern vorab digital übermittelt. Wurden alle Ausschlusskriterien mit nein beantwortet, wurde der Person anschließend ein Termin für die Testung bekannt gegeben und die Einverständniserklärung zugesandt.

Alle Probandinnen und Probanden nahmen freiwillig an der Studie teil.

### **5.3.5 Demographische Daten der Stichprobe**

An der Studie nahmen 40 Probandinnen und Probanden teil (28 Frauen, 12 Männer). Das durchschnittliche Alter betrug  $40.1 \pm 13.8$  Jahre (min. 18 Jahre und max. 81 Jahre).

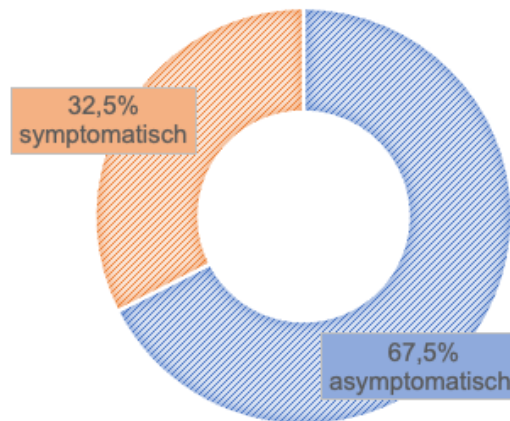
### **5.3.6 Daten zur Händigkeit**

Anhand eines Fragebogens, welcher von den Studienteilnehmerinnen und -teilnehmern vor Beginn der Testung vor Ort ausgefüllt wurde, lässt sich eine Aussage über die Verteilung der Händigkeit in der Stichprobe treffen. Von den 40 Teilnehmenden gaben vier an linkshändig zu sein (10%), 35 rechtshändig (87.5%) und eine Person beidhändig (2.5%).

### **5.3.7 Daten zu vorliegender Symptomatik**

Der im vorigen Kapitel erwähnte Fragebogen erhob neben der Händigkeit der Probandinnen und Probanden auch das Vorhandensein von möglichen Symptomen, die auf eine Dysfunktion der ersten Rippe hindeuten könnten. Der Fragebogen wurde basierend auf der in Kapitel 2.3.2 beschriebenen Symptomatik verfasst.

Die Aufteilung der Stichprobe in symptomatisch und asymptomatisch ist in nachfolgender Abbildung 1 dargestellt. An der Studie nahmen 27 asymptomatische (67.5%) und 13 symptomatische Probandinnen und Probanden (32.5%) teil.



**Abbildung 1.** Anteil der symptomatischen und asymptomatischen Studienteilnehmerinnen – und teilnehmer

Die zutreffenden Symptome in den Bereichen Schultergürtel, Armen und/oder Händen gliederten sich folgendermaßen auf: Sechs Probandinnen bzw. Probanden gaben an, in einem der oben genannten Bereiche, Schmerzen gehabt zu haben (15%). Bei einer Person verstärkten sich diese Schmerzen durch Tätigkeiten wie Heben oder Tragen (2.5%). Eine Schmerzverstärkung bei tiefer Ein- bzw. Ausatmung traf bei niemandem zu. Neun Probandinnen und Probanden gaben an, unter Kribbeln bzw. einschlafendem Gefühl zu leiden (22.5%) und zweimal wurde eine Kraftlosigkeit angegeben (5%). Das Symptom der Schwellung traf bei keinem Teilnehmenden zu.

## 5.4 Testerin und Tester

Wie in Kapitel 2.2.1 erklärt, bedarf es für die Ausführung einer Interrater Reliabilitätsstudie zweier Testerinnen bzw. Tester. In vorliegender Studie fungierte die Autorin als Testerin. Der Autor der Interrater Reliabilitätsstudie des costoclaviculären Manövers – Maximilian Puhr – wurde als zweiter Tester herangezogen. Beide Untersuchungspersonen gehen seit 2016 dem Beruf der Physiotherapie nach und verfügen somit über eine annähernd gleiche Berufserfahrung. Zusätzlich absolvieren beide seit 2016 eine Osteopathie-Ausbildung und befanden sich zum Zeitpunkt der Studie im Masterlehrgang für Osteopathie.

## 5.5 Studienleitung

Als Studienleitung stellte sich Christina Halasz, MSc (Physiotherapeutin und Osteopathin mit langjähriger Berufserfahrung) zur Verfügung. Die Studienleiterin war am Tag der Testung für folgende Tätigkeiten zuständig:

- Annahme der Einverständnisbögen
- Annahme der Fragebögen zur Abklärung der Ausschlusskriterien
- Annahme der Fragebögen bezüglich Händigkeit und Symptomatik

- Randomisierung der Probandinnen und Probanden (per Los) und somit Zuteilung zu den Testräumen
- Markierung der Probandinnen und Probanden mit einem nummerierten Aufkleber
- Markierung des Processus spinosus von Th1 mit einem Hautstift
- Aufsicht über das Testgeschehen – Kontrolle und Sicherstellung der Verblindung
- Ausgabe einer Tafel Schokolade an die Probandinnen und Probanden als Entschädigung für den durch die Testung entstandenen Zeitaufwand

## **5.6 Trainingsphase**

Vor Beginn der Trainingsphase einigte sich das Studienpersonal bezüglich einer standardisierten Durchführung in folgenden Punkten in Anlehnung an die Empfehlungen von Patijn (2019):

- Ausgangsposition der Testerin bzw. des Testers
- Ausgangsposition der Probandinnen bzw. der Probanden
- anatomische Orientierungspunkte zur Erleichterung der Palpation
- Handhaltung bzw. Grifftechnik
- Ausführung des Tests
- Anzahl der Wiederholungen
- verbale Instruktion der Probandinnen und Probanden
- mögliche Interpretationsspielräume bei der Beurteilung des Testverfahrens

In der gemeinsamen Trainingsphase wurde das vorab standardisierte Verfahren anhand von 10 Probandinnen und Probanden gemeinsam ausprobiert, um etwaige Ungereimtheiten in der Durchführung zu evaluieren.

Anschließend wurde der standardisierte Ablauf des Flexionstests von der Testerin und dem Tester im Zuge des Arbeitsalltags an so vielen Patientinnen und Patienten wie möglich durchgeführt.

Einen Monat vor der tatsächlichen Testung wurde zusätzlich ein Probedurchlauf absolviert, um einen konkreten Zeitplan für die Testung erstellen zu können. Erst nach Absolvierung der Trainingsphase wurde die Testung durchgeführt.

## **5.7 Ablauf der Testung**

Die Testung fand am 24. April 2021 in der Praxis Osteopathie Gröbming statt. Da es sich bei der Praxis um einen extramuralen Bereich handelt, konnte auf ein Ethikvotum verzichtet werden.

Für den Ablauf der Studie galten dieselben Datenschutzbestimmungen und Hygienestandards, wie sie auch in osteopathischen Praxen vorgeschrieben sind.

Alle Teilnehmerinnen und Teilnehmer sowie auch das Studienpersonal waren zum Tragen einer FFP2-Maske verpflichtet. Zudem wurden, sofern möglich, die Abstandsbestimmungen streng eingehalten.

Pro 20 Minuten wurden zwei Probandinnen bzw. Probanden geladen. Im Vorraum wurden die unterschriebenen Einverständniserklärungen sowie die ausgefüllten Fragebögen zur Abklärung der Ausschlusskriterien abgegeben. Die Fragebögen zur Erhebung der Händigkeit bzw. möglicher Symptomatik wurden von Seiten der Teilnehmerinnen und Teilnehmer vor Ort ausgefüllt. Um die Testerin bzw. den Tester nicht vorab in ihrer Testbeurteilung zu beeinflussen, wurden die Fragebögen an die unabhängige Studienleitung gereicht.

Den Probandinnen und Probanden wurden anschließend nummerierte Aufkleber auf den Oberschenkel geklebt.

Darauf traten die zu testenden Personen in zwei unterschiedliche Untersuchungsräume ein. Dort wurden sie von der Studienleiterin gebeten, das T-Shirt auszuziehen, außer der zu untersuchende Bereich war ohnehin nicht von Kleidung verdeckt. Die Studienleiterin markierte nun jeweils den Processus spinosus von Th1 mit einem Hautstift, um es der Testerin und dem Tester in weiterer Folge zu erleichtern, die erste Rippe rascher zu identifizieren.

Erst jetzt betraten die beiden Untersuchungspersonen die zwei Räume und notierten jeweils die Nummer der Probandin bzw. des Probanden in den Ergebnisbogen. Anschließend lasen die beiden einen standardisierten Text zur Erklärung der Testung vor, um etwaige individuelle Färbungen in der Anweisung ausschließen zu können. Dann wurden hintereinander beide Untersuchungen (Flexionstest und costoclaviculäres Manöver) an einer Probandin bzw. einem Probanden durchgeführt.

Da der Flexionstest eine rein palpatorische Untersuchung, gekoppelt mit einer aktiven Bewegung von Seiten der Probandin bzw. des Probanden, darstellt, waren im Zuge der Studie keinerlei Risiken bzw. potenzielle Schädigungen für die Teilnehmenden zu erwarten. Die Probandinnen und Probanden wurden aufgefordert, sich bei Schmerzen während der Testung jederzeit zu melden, damit umgehend abgebrochen werden kann. Dieser Fall trat jedoch bei keiner der 40 Testpersonen auf.

Die Ergebnisse wurden von der Testerin und dem Tester eigenständig auf einem Ergebnisbogen notiert, ohne der Probandin bzw. dem Probanden Einblick zu gewähren. Die Ergebnisbögen wurden verdeckt abgelegt, um die zweite Testperson nicht zu beeinflussen. Im Anschluss wechselten die Testerin und der Tester den Behandlungsraum und führten die Testungen im zweiten Raum abermals durch.

Anschließend verließen die Probandinnen und Probanden die Untersuchungsräume und bekamen von der Studienleitung eine Tafel Schokolade als Entschädigung für den durch die Testung entstandenen Zeitaufwand ausgehändigt.

Die Probandinnen und Probanden wurden angehalten sich telefonisch beim Studienpersonal zu melden falls wider Erwarten im Anschluss an die Testung Symptome oder Unwohlsein auftreten sollten.

Waren die Probandinnen und Probanden an den Ergebnissen interessiert, wurden die individuellen Daten nach Abschluss der Testung digital übermittelt.

Nach Ende der Testung fand eine Nachbesprechung des Studienpersonals statt. Hierbei wurden die Eindrücke bezüglich der Durchführung sowie etwaige Limitationen besprochen. Aspekte aus dieser Besprechung werden in Kapitel 7 aufgegriffen.

## **5.8 Verblindung**

Die Verblindung zwischen den beiden Untersuchenden erfolgte insofern, dass sich die beiden während der Testung nicht miteinander austauschen durften und somit keinerlei Einblick in die Ergebnisse des absolvierten Tests der anderen Person erlangen konnten.

Die Verblindung zwischen der Testerin bzw. dem Tester und den Probandinnen bzw. Probanden wurde durch das Unterbinden von verbalem Austausch bezüglich möglicher vorhandener Symptomatik sowie Ergebnissen der Testung gewährleistet.

## **5.9 Randomisierung**

Die Randomisierung der Probandinnen und Probanden erfolgte beim Betreten der Praxis mittels Los. Somit wurde entschieden, wer in welchem Raum und somit bei wem (Testerin oder Tester) beginnt.

Da nicht nur ein Testverfahren auf seine Reliabilität hin geprüft wurde, sondern auch ein zweites Testverfahren, stellte sich zudem die Frage der Reihenfolge der Untersuchungen. Beide Testungen untersuchen den Bereich der oberen Thoraxapertur, weswegen eine Beeinflussung des zuerst durchgeführten Tests auf den zweiten Test vermieden werden sollte.

Aus diesem Grund wurden unterschiedlich randomisierte Testreihenfolgen angewandt:

- Bei jeder Probandin bzw. jedem Probanden wurde einmal mit dem Flexionstest und einmal mit dem Test für das costoclaviculäre Manöver begonnen.
- Zusätzlich wechselten die Testerin und der Tester die Testreihenfolge, sodass jeder 20 mal mit dem Flexionstest und 20 mal mit dem Test für das costoclaviculäre Manöver startete.

## **5.10 Markierung des Processus spinosus von Th1**

In Reviews über die Reliabilität von manuellen Untersuchungen und Palpationstechniken der Wirbelsäule wird häufig die Wichtigkeit einer Markierung von ossären Orientierungspunkten durch eine unabhängige Person betont (Haneline & Young, 2009; Hollerwöger, 2006).

Demnach erschien es für die Qualität vorliegender Studie essenziell, vorab anatomische Orientierungspunkte zur Palpation der ersten Rippe an den Probandinnen und Probanden zu markieren.

In der manuellen Medizin werden unterschiedliche Herangehensweisen für die Palpation der ersten Rippe beschrieben. Laekeman und Kreutzer (2009, S. 293) empfehlen die Palpation durch den Musculus Trapezius hindurch. Über eine ipsilaterale Lateralflexion und eine kontralaterale Rotation der Halswirbelsäule kann eine Annäherung des Musculus Trapezius erreicht werden, was die Palpation der ersten Rippe erleichtert. Palpiert wird nach einer knöchernen Struktur – der ersten Rippe.

Krumholz (1998) nimmt als Orientierung für die erste Rippe den Processus spinosus von Th1 als Ausgangspunkt. Die erste Rippe ist hierbei vier bis fünf Querfinger neben dem Processus spinosus und knapp oberhalb davon zu finden.

Loyd et al. (2014) untersuchten in ihrer Studie an Präparaten, welche anatomischen Orientierungspunkte sich für die Palpation der ersten Rippe als praktikabel erweisen. Die Autorinnen und Autoren kamen zu dem Schluss, dass der Abstand zwischen beiden Processi mastoidei gleich breit wie der Abstand zwischen den beiden medialen Enden der ersten Rippen ist. Reichert (2018, S. 342f) schlussfolgert daraus folgende Projektionshilfe für die erste Rippe: Wird ausgehend von den Processi mastoidei eine vertikale Falllinie nach unten gezogen, trifft der Untersuchende auf die erste Rippe. Die Orientierung bezüglich der Höhe kann über den Processus spinosus von Th1 ermittelt werden. Einen Querfinger höher findet sich der dazugehörige Processus transversus und daneben die erste Rippe.

In vorliegender Studie wurden alle drei oben genannten Herangehensweisen in folgender Reihenfolge angewandt:

- Orientierung der Höhe: ausgehend von dem Processus spinosus von Th1 einen Querfinger nach cranial
- Orientierung der Breite: über die vertikale Falllinie von beiden Processi mastoidei
- Bestätigung: über die Palpation einer knöchernen Struktur unter den Daumen

Um diesem Palpations-Leitfaden für die erste Rippe folgen zu können, war die Markierung des Processus spinosus von Th1 unumgänglich. Die Markierung wurde von der Studienleiterin mit einem Hautstift durchgeführt. Die Studienleiterin folgte für die Palpation des Processus spinosus von Th1 folgender Anleitung von Reichert (2018, 338ff): Bei passiver Extension der Halswirbelsäule bewegt der Processus spinosus von C6 Richtung Extension mit, während der Processus spinosus von C7 in Relation unbewegt bleibt. Einen Processus spinosus darunter

befindet sich der Processus spinosus von Th1. Zur Verifizierung von Th1 wird im Anschluss eine passive Rotations-/Lateralflexionsbewegung der Halswirbelsäule eingeleitet. Hierbei sollte ein Mitbewegen des Processus spinosus von C7 gut spürbar sein, während der Processus spinosus von Th1 unbewegt bleibt bzw. deutlich weniger stark als C7 mitbewegt.

## 5.11 Durchführung des Flexionstests

Als Ausgangsstellung für die Testung der Mobilität der ersten Rippe wurde der angelehnte Sitz gewählt. In der ursprünglichen Testbeschreibung von Peeters und Lason (2015, S. 79) wird der entspannte Sitz beschrieben, welcher jedoch stark der subjektiven Wahrnehmung der Einzelnen bzw. des Einzelnen unterliegt und somit schwer reproduzierbar ist. Im Zuge vorliegender Interrater Reliabilitätsstudie ist eine standardisierte Ausgangsposition von hoher Wichtigkeit – daher wurde der angelehnte Sitz gewählt.

Die Testerin bzw. der Tester orientierten sich am markierten Processus spinosus von Th1 und suchten mit den Daumen, wie im vorhergehenden Kapitel beschrieben, beidseits den Angulus costae der ersten Rippen auf. Anschließend wurden die Probandinnen und Probanden aufgefordert, den Kopf langsam zu flektieren. In der Testbeschreibung von Peeters und Lason (2015, S. 79) gibt es keinerlei Angaben zum Ausmaß der notwendigen Flexion. Um eine normierte Endposition für eine bessere Vergleichbarkeit der Durchführung zu erhalten, wurden die Probandinnen und Probanden instruiert, die Halswirbelsäule so weit wie möglich zu flektieren (siehe Abbildung 2).



**Abbildung 2.** Endposition des Flexionstests

Die Daumen der Testerin bzw. des Testers folgten der Bewegung der ersten Rippen nach cranial. Bewegte eine Seite schneller als die andere, wurde ebendiese Seite als auffällig gewertet. Somit standen den Untersuchenden folgende Beurteilungskriterien zur Verfügung:

- unauffällig
- links auffällig
- rechts auffällig

Das Ergebnis der Testung wurde anschließend in den Beurteilungsbogen eingetragen.

Ein mehrmaliges Testen kann, wie in Kapitel 2.2.1 beschrieben, zu einem mobilisierenden Effekt führen, was folglich einen Einfluss auf die Testung durch die zweite Untersuchungsperson gehabt hätte. Daher durfte der Test nur einmal durchgeführt werden. Ausgenommen davon waren Fälle, in denen die Probandin bzw. der Proband den Kopf für eine konkrete Beurteilung nicht ausreichend weit bzw. zu schnell bewegte.

## **5.12 Statistische Datenaufbereitung**

Die erhobenen Daten wurden in den Datenverarbeitungsprogrammen SPSS und Microsoft Excel aufbereitet.

### **5.12.1 Darstellung der Daten**

Die Daten werden, wie von Patijn (2019) empfohlen, in einer Kreuztabelle dargestellt.

Tabelle 2 gibt folgende Auskunft: T1 und T2 stehen für das Testpersonal, also die Testerin bzw. den Tester. Die Zahlen 0, 1 und 2 stehen für die Beurteilungskategorien des Testverfahrens. In vorliegender Studie steht die Kategorie 0 für ein unauffälliges Testergebnis, Kategorie 1 für eine Auffälligkeit links und Kategorie 2 für eine Auffälligkeit rechts.

Die Variablen a bis i repräsentieren die beobachteten absoluten Häufigkeiten der Testergebnisse. Die hellblau hinterlegten Zellen stellen die Nicht-Übereinstimmungen dar, die dunkelblauen Zellen mit den fett gedruckten Variablen a, e und i in der Diagonale, zeigen die Übereinstimmungen der Testerin bzw. des Testers auf. Die Spaltensummen und Zeilensummen stellen die Summen der absoluten Häufigkeiten dar und ergeben in ihrer Summe die Variable n (= Anzahl der getesteten Probandinnen und Probanden) (Döring & Bortz, 2016, S. 567f; Schäfer & Schöttker-Königer, 2015, S. 136f).

**Tabelle 2.** Kreuztabelle – Quelle: adaptiert nach Schäfer und Schöttker-Königer (2015, S. 137)

		T1			Σ
		Kategorie	0	1	
T2	0	a	b	c	a+b+c
	1	d	e	f	d+e+f
	2	g	h	i	g+h+i
Σ		a+d+g	b+e+h	c+f+i	n

### 5.12.2 Berechnung und Interpretation des Cohens Kappa Koeffizienten

Zur Berechnung des Cohens Kappa Koeffizienten bedarf es der vorhergehenden Berechnung folgender Variablen (Döring & Bortz, 2016, S. 567f; Patijn, 2019):

- $p_o$  beschreibt die prozentuelle Übereinstimmung
- $p_e$  beschreibt die prozentuelle Zufallsübereinstimmung

Die Formel für den Cohens Kappa Koeffizienten lautet wie folgt (Bortz & Lienert, 2008, S. 311):

$$\kappa = \frac{p_o - p_e}{1 - p_e}$$

Aus den erhobenen Daten wird ein primärer Cohens Kappa Koeffizient errechnet – in diesem sind Werte von allen Studienteilnehmerinnen und -teilnehmern enthalten (mit drei Ergebniskategorien: unauffällig, links auffällig und rechts ausfällig).

Zusätzlich werden drei weitere Kappa Koeffizienten errechnet:

- ein Wert für alle Studienteilnehmerinnen und -teilnehmer (mit nur zwei Ergebniskategorien: unauffällig oder auffällig – diesmal ohne Berücksichtigung der Seiten)
- je zwei Werte, geschlechtlich getrennt jeweils für Probandinnen und für Probanden, um etwaige geschlechtsspezifische Unterschiede sichtbar zu machen

Für jeden Cohens Kappa Koeffizienten wird das dazugehörige 95% Konfidenzintervall (= 95% KI) kalkuliert.

Die Koeffizienten können Werte zwischen -1 und +1 annehmen (Patijn, 2019) und sollen mithilfe der von Landis und Koch (1977, S. 165) beschriebenen Klassifikation interpretiert werden (siehe Tabelle 3).

**Tabelle 3.** Klassifikation der Cohens Kappa Koeffizienten – Quelle: adaptiert nach Landis und Koch (1977, S. 165)

<b>Cohens Kappa Koeffizient <math>\kappa</math></b>	<b>Grad der Übereinstimmung</b>
<0.00	poor (mangelhaft)
0.00 – 0.20	slight (gering)
0.21 – 0.40	fair (mäßig)
0.41 – 0.60	moderate (mittelmäßig)
0.61 – 0.80	substantial (beachtlich)
0.81 – 1.00	almost perfect (fast vollkommen)

### **5.12.3 Berechnung der Häufigkeitsverteilungen der Testergebnisse**

Aus den erhobenen Daten sollen Häufigkeitsverteilungen für beide Untersuchungspersonen getrennt mittels absoluter und prozentueller Häufigkeiten berechnet werden.

Neben der Verteilung der gefundenen Dysfunktionen auf die einzelnen Kategorien (0, 1 und 2) soll auch die Geschlechterverteilung (Probandinnen und Probanden) dargestellt werden.

### **5.12.4 Beschreibung des Zusammenhangs zwischen Ergebnissen und Händigkeit**

Die erhobenen Ergebnisse, insbesondere die tatsächlichen Übereinstimmungen beider Untersuchungspersonen, sollen in Verbindung mit der Händigkeit der Probandinnen und Probanden gestellt werden.

## 6 Ergebnisse

In den folgenden Kapiteln werden die Ergebnisse der Studie dargestellt. Auf jeden errechneten Cohens Kappa Koeffizienten wird in einem eigenständigen Kapitel eingegangen.

### 6.1 Kappa Koeffizient mit drei Ergebniskategorien

In untenstehender Tabelle 4 sind die beobachteten Häufigkeiten der Testbeurteilung aller Probandinnen und Probanden gelistet. Die Kategorie 0 steht für unauffällig, Kategorie 1 für links auffällig und Kategorie 2 für rechts auffällig.

Absolut gesehen gibt es 13 Übereinstimmungen der beiden Untersuchenden, die sich folgendermaßen auf die drei Ergebniskategorien aufgliedern: Kategorie 0 weist sechs Übereinstimmungen auf, Kategorie 1 zwei und Kategorie 2 fünf Übereinstimmungen. Daraus ergibt sich eine prozentuelle Übereinstimmung von  $p_o = 32.5\%$ . Die prozentuelle Zufallsübereinstimmung  $p_e$  liegt bei 29.6%. Der p-Wert beträgt  $p = 0.66$ .

Der aus den vorliegenden Daten errechnete Kappa Koeffizient ergibt somit  $\kappa = 0.04$  und das 95% KI erstreckt sich von -0.15 bis 0.23.

**Tabelle 4.** Kreuztabelle – alle Probandinnen und Probanden – drei Ergebniskategorien

		T1			Σ
		Kategorie	0	1	
T2	0	6	5	5	16
	1	0	2	1	3
	2	6	10	5	21
Σ		12	17	11	40

### 6.2 Kappa Koeffizient mit zwei Ergebniskategorien

Nachfolgender Tabelle 5 sind die beobachteten Häufigkeiten der Testbeurteilung von den Untersuchenden bei der Betrachtung mit zwei Ergebniskategorien (unauffällig oder auffällig) ohne Berücksichtigung der Seiten zu entnehmen. Die Kategorie 0 steht in diesem Fall für unauffällig, Kategorie 1 für auffällig.

Bei dieser Darstellung sind 24 Übereinstimmungen der Testerin bzw. des Testers zu verzeichnen. Kategorie 0 weist hier sechs Übereinstimmungen auf und Kategorie 1 zählt 18 Übereinstimmungen. Prozentuell gesehen ergibt das eine Übereinstimmung von  $p_o = 60\%$ . Die zufällige Übereinstimmung  $p_e$  liegt bei 54%. Der p-Wert beträgt  $p = 0.40$ .

Der Kappa Koeffizient ergibt somit  $\kappa = 0.13$  und das 95% KI erstreckt sich von -0.18 bis 0.44.

**Tabelle 5.** Kreuztabelle – alle Probandinnen und Probanden – zwei Ergebniskategorien

		T1		Σ
		Kategorie 0	Kategorie 1	
T2	0	6	10	16
	1	6	18	24
Σ		12	28	40

### 6.3 Kappa Koeffizient mit drei Ergebniskategorien für Probandinnen

In Tabelle 6 sind die beobachteten Häufigkeiten der Testbeurteilung für Probandinnen gelistet. Bei der Berechnung des Kappa Koeffizienten für weibliche Studienteilnehmende kamen die Testerin und der Tester auf neun Übereinstimmungen, welche sich folgendermaßen auf die drei Ergebniskategorien aufgliedern: Kategorie 0 weist drei Übereinstimmungen auf, Kategorie 1 zwei und Kategorie 2 vier Übereinstimmungen. Prozentuell gesehen ergibt das eine Übereinstimmung von  $p_o = 32.1\%$ . Die Übereinstimmung durch Zufall liegt bei  $p_e = 30.1\%$ . Der p-Wert beträgt  $p = 0.80$ .

Der Kappa Koeffizient ergibt somit  $\kappa = 0.03$  und das 95% KI erstreckt sich von -0.20 bis 0.26.

**Tabelle 6.** Kreuztabelle – Probandinnen – drei Ergebniskategorien

		T1			Σ
		Kategorie 0	Kategorie 1	Kategorie 2	
T2	0	3	4	3	10
	1	0	2	1	3
	2	5	6	4	15
Σ		8	12	8	28

### 6.4 Kappa Koeffizient mit drei Ergebniskategorien für Probanden

In Tabelle 7 sind die beobachteten Häufigkeiten der Testbeurteilung für Probanden gelistet. Bei der Beurteilung der männlichen Studienteilnehmer stimmen die Testerin und der Tester in vier Fällen überein, was sich folgendermaßen auf die drei Ergebniskategorien aufgliedern lässt: Kategorie 0 weist drei Übereinstimmungen auf, in Kategorie 1 liegt keine einzige Übereinstimmung vor und in Kategorie 2 wurde eine Übereinstimmung gezählt.

Prozentuell gesehen ergibt das eine Übereinstimmung von  $p_o = 33.3\%$ . Die zufällige Übereinstimmung  $p_e$  liegt bei  $29.2\%$ . Der p-Wert beträgt  $p = 0.70$ .

Der Kappa Koeffizient ergibt somit  $\kappa = 0.06$  und das 95% KI erstreckt sich von -0.24 bis 0.36.

**Tabelle 7.** Kreuztabelle – Probanden – drei Ergebniskategorien

		T1			Σ
		Kategorie 0	Kategorie 1	Kategorie 2	
T2	0	3	1	2	6
	1	0	0	0	0
	2	1	4	1	6
Σ		4	5	3	12

## 6.5 Übersicht über alle errechneten Kappa Koeffizienten

Für eine bessere Übersicht der errechneten Cohens Kappa Koeffizienten werden die vier Werte mit ihren 95% KI in untenstehender Tabelle 8 zusammenfassend dargestellt. Die errechneten Kappa Koeffizienten befinden sich zwischen  $\kappa = 0.03$  und  $\kappa = 0.13$ . Der niedrigste errechnete Wert wurde bei der Berechnung der weiblichen Teilnehmenden erreicht, der höchste errechnete Wert bei der Betrachtung mit zwei Ergebniskategorien.

**Tabelle 8.** Übersicht über alle errechneten Cohens Kappa Koeffizienten mit zugehörigem 95% KI

Stichprobe (n) – Anzahl der Kategorien	$\kappa$	95% KI Untergrenze	95% KI Obergrenze
gesamte Stichprobe (n = 40) – 3 Kategorien	0.04	-0.15	0.23
gesamte Stichprobe (n = 40) – 2 Kategorien	0.13	-0.18	0.44
Probandinnen (n = 28) – 3 Kategorien	0.03	-0.20	0.26
Probanden (n = 12) – 3 Kategorien	0.06	-0.24	0.36

## 6.6 Häufigkeitsverteilung der Testergebnisse

Werden die Ergebnisse der beiden Untersuchenden getrennt begutachtet, lassen sich unterschiedliche Häufigkeiten der gewählten Ergebniskategorien erkennen. Die Zahlen 0, 1 und 2 in Tabelle 9 stehen für die Ergebniskategorien des Flexionstests (0 = unauffällig, 1 = links auffällig, 2 = rechts auffällig).

**Tabelle 9.** Häufigkeitsverteilung der Testergebnisse

Kategorie	T 1		T 2	
	<i>absolut</i>	<i>prozentuell</i>	<i>absolut</i>	<i>prozentuell</i>
0 (unauffällig)	12	30%	16	40%
1 (links auffällig)	17	42.5%	3	7.5%
2 (rechts auffällig)	11	27.5%	21	52.5%

Der Tabelle kann folgendes entnommen werden:

Die Ergebniskategorie 0 wurde von beiden Untersuchenden am zweithäufigsten ausgewählt – von T1 zu 30% und von T2 zu 40%.

T1 entschied sich zu 42.5% und somit am häufigsten für die Kategorie 1. Am seltensten (27.5%) wurde Kategorie 2 gewählt. Bei T1 lässt sich also eine annähernd gleichmäßige Aufteilung auf die einzelnen Kategorien finden.

Bei T2 hingegen ist eine ungleiche Verteilung zugunsten der rechten Seite erkennbar – bei mehr als der Hälfte (52.5%) aller Probandinnen und Probanden wurde die Kategorie 2 gewählt. Die linke Seite (Kategorie 1) wählte T2 hingegen am seltensten (7.5%).

Wird das Ergebnis hinsichtlich zweier Ergebniskategorien, wie in Kapitel 5.12.2 beschrieben, betrachtet – unauffällig (0) und auffällig (1 und 2 gesamt) – hat T1 zu 70% eine Auffälligkeit gefunden, T2 hingegen zu 60%.

Tabelle 10 thematisiert die Verteilung in Hinblick auf das Geschlecht der Teilnehmenden. Die absoluten und prozentuellen Werte stellen die auffälligen Testergebnisse dar (Ergebniskategorie 1 und 2 gesamt). Hierbei wird ersichtlich, dass beide Untersuchende mehr Auffälligkeiten bei den Probandinnen gefunden haben als bei den Probanden.

**Tabelle 10.** Geschlechtsverteilung der auffälligen Testergebnisse

Geschlecht	T 1		T 2	
	<i>absolut</i>	<i>prozentuell</i>	<i>absolut</i>	<i>prozentuell</i>
♀	20	71%	18	64%
♂	8	67%	6	50%

## 6.7 Zusammenhang zwischen Ergebnissen und Händigkeit

In den insgesamt 13 gefundenen Übereinstimmungen der Testerin und des Testers befinden sich sieben Übereinstimmungen für eine auffällige Seite (zweimal links und fünfmal rechts). In drei Fällen davon deckt sich das Ergebnis (auffällige Seite) mit der Händigkeit der Probandin bzw. des Probanden. So wurde bei zwei Teilnehmerinnen bzw. Teilnehmern mit Rechtshändigkeit auch eine auffällige rechte Seite gefunden. Bei einer Teilnehmerin mit Beidhändigkeit konnte ebenfalls von beiden Untersuchenden eine auffällige rechte Seite identifiziert werden.

## 6.8 Überprüfung der Hypothesen

Die in Kapitel 4.2 beschriebenen Hypothesen können aufgrund der erhobenen Daten verifiziert bzw. falsifiziert werden. Zur Überprüfung der Hypothesen wurde der Cohens Kappa Koeffizient mit drei Ergebniskategorien für alle Probandinnen und Probanden herangezogen.

Nullhypothese  $H_0$ : Die Übereinstimmung der Ergebnisse des Flexionstests für die erste Rippe ist nicht substanziell (Landis & Koch, 1977, S. 165), der Cohens Kappa Koeffizient  $\kappa$  ist somit kleiner als 0.61.

→ Der Kappa Koeffizient liegt bei  $\kappa = 0.04$  und ist somit kleiner als 0.61. Das dazugehörige 95% KI liegt zwischen -0.15 und 0.23. Daher wird die Nullhypothese angenommen und besagt, dass die Ergebnisse von den einzelnen Untersuchungspersonen abhängig sind.

Alternativhypothese  $H_1$ : Die Übereinstimmung der Ergebnisse des Flexionstests für die erste Rippe ist mindestens substanziell (Landis & Koch, 1977, S. 165), der Cohens Kappa Koeffizient  $\kappa$  ist somit größer bzw. gleich 0.61.

→ Der Kappa Koeffizient liegt bei  $\kappa = 0.04$ . Das dazugehörige 95% KI liegt zwischen -0.15 und 0.23. Daher wird die Alternativhypothese abgelehnt.

## 7 Diskussion

Zur besseren Übersicht gliedert sich dieses Kapitel in mehrere Unterkapitel auf. Die Einbettung der Ergebnisse in die bestehende Forschungslandschaft, die Interpretation der Ergebnisse, die Bedeutung für die Osteopathie sowie die Limitationen der Studie werden separat diskutiert.

### 7.1 Vergleich der Ergebnisse mit bestehender Literatur

Die Bandbreite an Cohens Kappa Koeffizienten aus bisher publizierten Studien (siehe Kapitel 3.1), welche ebenfalls eine Testung für die Mobilität der ersten Rippe auf ihre Verlässlichkeit untersuchten, reicht von  $\kappa = 0.06$  (Fjellner et al., 1999) bis hin zu  $\kappa = 1.0$  (Lindgren et al., 1992). Aus Gründen der besseren Vergleichbarkeit mit Werten, die aus der bisherigen Forschung generiert wurden, wird ausschließlich der Wert mit drei Ergebniskategorien für alle Studienteilnehmerinnen und -teilnehmer herangezogen.

Der im Zuge dieser Studie erhobene Kappa Koeffizient für die gesamte Stichprobe von  $\kappa = 0.04$  stellt den niedrigsten Wert für eine Untersuchung der Mobilität der ersten Rippe im Vergleich mit der Literatur dar. In den anschließenden Kapiteln sollen mögliche, für diesen niedrigen Wert verantwortliche, Ursachen angeführt werden.

### 7.2 Interpretation der Ergebnisse

Die Analyse und Interpretation der aus den Daten errechneten Cohens Kappa Koeffizienten in den anschließenden Kapiteln stützt sich auf die Klassifikation von Landis und Koch (1977, S. 165) (siehe Kapitel 5.12.2).

#### 7.2.1 Interpretation des Kappa Koeffizienten mit drei Ergebniskategorien

Der errechnete Cohens Kappa Koeffizient für alle Probandinnen und Probanden ergibt  $\kappa = 0.04$ . Dieser Wert wird der Kategorie der geringen Übereinstimmung zugeteilt. Das dazu errechnete 95% KI erstreckt sich von -0.15 bis 0.23 und besagt, dass bei zukünftigen Studien der Kappa Koeffizient mit 95-prozentiger Wahrscheinlichkeit in besagtem Wertebereich liegen wird.

Würde ein potenziell zukünftiger Wert an der Untergrenze dieses 95% KI liegen, könnte es sich auch um einen negativen Kappa Koeffizienten handeln. Das würde bedeuten, die prozentuelle Übereinstimmung  $p_o$  liegt sogar unter der zufällig entstandenen Übereinstimmung  $p_e$  (Bortz & Lienert, 2008, S. 311). Negative Kappa Werte entsprechen einer mangelhaften Übereinstimmung.

Ein potenziell zukünftiger Kappa Koeffizient könnte hingegen auch an der Obergrenze des 95% KI angesiedelt sein und würde bereits der Kategorie der mäßigen Übereinstimmung zugeordnet werden.

Demzufolge reicht die Bandbreite an potenziellen Kappa Koeffizienten in zukünftigen Studien von mangelhafter bis mäßiger Übereinstimmung.

Über die Ursache für die geringe Übereinstimmung zwischen der Testerin und dem Tester kann nur spekuliert werden. Folgendes könnte dafür verantwortlich sein:

Vorliegende Studie wurde, wie in Kapitel 5.2 erwähnt, in Kooperation mit einer zweiten Interrater Reliabilitätsstudie durchgeführt. Dies könnte eine unbemerkte gegenseitige Beeinflussung der Tests mit sich gebracht haben. Möglicherweise wurde von Seiten der Testerin bzw. des Testers nach einer rechtsseitigen Auffälligkeit beim costoclaviculären Manöver unbewusst auch ein rechtsseitig auffälliges Ergebnis des Flexionstest erwartet und vice versa. In vorliegender Arbeit wurde versucht mittels randomisierter Testreihenfolge dieser Beeinflussung entgegenzusteuern. Nichtsdestotrotz könnte sich dieses Phänomen möglicherweise negativ auf die Ergebnisse der Studie ausgewirkt haben und so den niedrigen Wert der gefundenen Übereinstimmungen erklären.

Einen weiteren möglichen Einfluss auf die stark voneinander abweichenden Ergebnisse der Testerin und des Testers könnte die Ausgangsposition der Probandinnen und Probanden verursacht haben. Sowohl die Testerin als auch der Tester gaben in der Nachbesprechung der Untersuchung an, bei manchen Probandinnen und Probanden eine Form von Anspannung bzw. Nervosität bemerkt zu haben. Dies drückte sich oftmals in einer angespannten Haltung aus (z.B. leicht hochgezogener Schultergürtel und somit angespannter Musculus trapezius pars descendens). Bei der Untersuchung durch die zweite Person wusste die Probandin bzw. der Proband bereits, was auf sie bzw. ihn zukommt, was sich folglich in einer entspannteren Grundhaltung äußerte. Somit könnte es trotz der Aufforderung in der Test-Instruktion, sich entspannt hinzusetzen, geringfügige Unterschiede in der Ausgangsposition der Probandinnen und Probanden gegeben haben. Das wiederum könnte eine unterschiedliche Ausgangssituation für den applizierten Test mit sich gebracht haben und somit einen ungewünschten Einfluss auf die Testergebnisse gehabt haben.

Eine weitere mögliche Ursache für die niedrige Übereinstimmung könnte die Beeinflussung durch optische Auffälligkeiten an der oberen Extremität der Probandinnen und Probanden dargestellt haben. Die Testerin und der Tester gaben unmittelbar nach der Testung an, dass sie in einem konkreten Fall von einer Narbe an der oberen Extremität eines Probanden möglicherweise indirekt beeinflusst wurden. Überdies nahmen die beiden vor der Testung Auffälligkeiten wie Haltungsasymmetrien (z.B. Schultergürtel in Elevation oder Protraktion) an den Probandinnen und Probanden wahr. Dieses Verhalten entspricht dem Arbeiten im

klinischen Alltag, wo die Inspektion einen wichtigen Teil der osteopathischen Untersuchung darstellt. Möglicherweise ließen die Testerin und der Tester das Beobachtete unbewusst in ihre Beurteilung der Tests miteinfließen.

Sowohl die Studienleitung als auch die beiden Testpersonen gaben an, in der Palpation der ossären Strukturen bei einigen Probandinnen und Probanden Schwierigkeiten gehabt zu haben. So stellte sich für die Studienleitung das Auffinden des Processus spinosus von Th1 als nicht immer eindeutig heraus und für die Testerin und den Tester das Auffinden der ersten Rippe. Die beiden Untersuchenden waren zudem von der Markierung der richtigen Struktur abhängig (in diesem Fall war dies der Processus spinosus von Th1). Bei der Markierung eines anderen Wirbelsäulensegments durch die Studienleiterin (zum Beispiel ein Segment höher oder tiefer), veränderte sich die Aussage des Testverfahrens dahingehend, dass womöglich nicht mehr die erste Rippe, sondern eine andere Struktur auf ihre Mobilität hin beurteilt wurde. Diese Beurteilung durch die Testerin bzw. den Tester erfolgte allerdings ohne deren Wissen darüber, da die beiden ja in der Annahme, die Markierung sei auf dem Processus spinosus von Th1, die Testung durchführten.

Die Schwierigkeiten in der Palpation traten laut dem Studienpersonal vor allem dann auf, wenn die Probandinnen und Probanden über einen muskulären Hypertonus des Musculus trapezius pars descendens verfügten bzw. auch wenn vermehrt subkutanes Fettgewebe den zu untersuchenden Bereich überdeckte.

Einen möglichen Einfluss auf die Palpation könnten außerdem unterschiedliche Body Mass Index Werte (= BMI-Werte) der Probandinnen und Probanden mit sich gebracht haben. Póvoa et al. (2018) bestätigen in ihrer Studie über die Lokalisation des Processus spinosus von C7, dass BMI-Werte > 25 mit einer niedrigeren Präzision bezüglich der Palpation einhergehen. Robinson et al. (2009) hingegen fanden in ihrer Studie über die Reliabilität und Validität der Palpation des Processus spinosus von C7 keinen Einfluss von erhöhten BMI-Werten der Probandinnen und Probanden auf die Präzision der Palpation. In vorliegender Studie wurde der BMI-Wert der Teilnehmerinnen und Teilnehmer nicht erhoben, wodurch ein etwaiger Einfluss auf die Palpation der ossären Strukturen und weiterführend auf die Studienergebnisse nicht sichtbar gemacht werden kann.

Durch die soeben beschriebenen Schwierigkeiten in der Palpation der ossären Strukturen, könnte es Unterschiede in der Testdurchführung gegeben haben, was wiederum eine Erklärung für die niedrige Übereinstimmung der Untersuchenden sein könnte.

Da die standardisierte Testbeschreibung die Vorgabe enthielt, nur einmal testen zu dürfen, mussten sich die beiden Untersuchenden nach einmaliger Testung für ein Kriterium entscheiden, auch wenn sie sich nicht hundertprozentig sicher waren. Sowohl die Testerin als

auch der Tester gaben in der Nachbesprechung an, dass sie bei manchen Probandinnen und Probanden ein zweites Mal hätten nachtesten wollen. Dies könnte auch eine Erklärung für die stark voneinander abweichenden Ergebnisse der Testerin und des Testers sein.

Für die Bewertung des Flexionstests standen den Untersuchungspersonen drei Beurteilungskriterien zur Verfügung (unauffällig, links auffällig, rechts auffällig). In der gemeinsamen Trainingsphase wurde der Interpretationsspielraum der Ergebnisse zwar besprochen, nichtsdestotrotz erwies sich die Beurteilung für die beiden als nicht immer eindeutig. Oftmals stellte sich die Frage, ab wann ist der Test auffällig und was ist noch in der Norm. Dieser, trotz Standardisierung vorhandene, subjektive Interpretationsspielraum der Untersuchenden könnte mitunter ebenso für die niedrige Übereinstimmung verantwortlich sein.

### **7.2.2 Interpretation des Kappa Koeffizienten mit zwei Ergebniskategorien**

Der Cohens Kappa Koeffizient für zwei Ergebniskategorien ohne Berücksichtigung der Seiten liegt bei  $\kappa = 0.13$  und wird in die Kategorie der geringen Übereinstimmung eingeordnet. Die Untergrenze des 95% KI liegt bei -0.18 und würde einer mangelhaften Übereinstimmung entsprechen. Die Obergrenze des 95% KI liegt bei 0.44, was einer moderaten Übereinstimmung gleichkäme. Zukünftige Kappa Werte könnten daher in vier unterschiedliche Kategorien (mangelhafte, geringe, mäßige und moderate Übereinstimmung) eingeteilt werden. Der für diese Betrachtung errechnete Kappa Wert von  $\kappa = 0.13$  stellt zwar mit  $p_o = 60\%$  die höchste Übereinstimmung aller errechneten Werte dieser Studie dar, ist jedoch aus folgenden Gründen kritisch zu hinterfragen: Der Flexionstest zielt nicht nur auf die Beurteilung auffällig bzw. unauffällig ab, sondern möchte vielmehr eine auffällige Seite identifizieren (siehe Berechnung mit drei Ergebniskategorien). Somit kommt diesem Kappa Wert kaum klinische Relevanz zu.

### **7.2.3 Interpretation des Kappa Koeffizienten mit drei Ergebniskategorien – geschlechterspezifisch**

Der errechnete Kappa Koeffizient für die weiblichen Studienteilnehmerinnen liegt bei  $\kappa = 0.03$  und wird demnach der Kategorie der geringen Übereinstimmung zugeordnet. Die Untergrenze des 95% KI liegt mit -0.20 im negativen Bereich, was somit einer mangelhaften Übereinstimmung gleichkäme. Mit der Obergrenze des 95% KI von 0.26 kann in zukünftigen Studien mit 95-prozentiger Wahrscheinlichkeit maximal ein Kappa Koeffizient mit mäßiger Übereinstimmung erreicht werden.

Die getrennte Berechnung für Probandinnen und Probanden zeigt, dass der errechnete Kappa Koeffizient für die männlichen Studienteilnehmer bei  $\kappa = 0.06$  liegt und somit doppelt so hoch wie bei den weiblichen Studienteilnehmerinnen ist. Somit wird dieser Wert der Kategorie der geringen Übereinstimmung zugeordnet. Ein potenzieller Kappa Koeffizient an der Untergrenze des 95% KI (-0.24) würde als mangelhafte Übereinstimmung interpretiert werden. Die Obergrenze des 95% KI liegt bei 0.36, was einer mäßigen Übereinstimmung entsprechen würde. Die Obergrenze liegt jedoch knapp an der Grenze (0.41) zur nächsthöheren Kategorie (moderate Übereinstimmung).

Eine mögliche Ursache für die, wenn auch nur minimalen, geschlechtsspezifischen Unterschiede der Ergebnisse könnte das Vorhandensein von vermehrtem subkutanem Fettgewebe darstellen. Laut Rieder und Lohff (2008, S. 131) verfügen Frauen über einen höheren Körperfettanteil (20-30%) als Männer (10-20%). Dies deckt sich auch mit den Beobachtungen der Testerin und des Testers während der Studie. Beide gaben im Anschluss an die Untersuchung an, bei den männlichen Teilnehmern besser gespürt zu haben, als bei den weiblichen. Die Vermutung der Testerin bzw. des Testers war, dass dies möglicherweise an dem vermehrten subkutanen Fettgewebe liegen könnte. Diese Erfahrung spiegelt sich auch in den Ergebnissen wider: Bei den Männern gab es mit  $p_o = 33,3\%$  eine dezent höhere Übereinstimmung als bei den Frauen mit  $p_o = 32,1\%$ .

#### **7.2.4 Interpretation der Häufigkeitsverteilung gefundener Testergebnisse**

Über die Ursachen der deutlichen Seitenunterschiede lässt sich nur spekulieren. Eine mögliche Erklärung für die Seitendifferenzen könnte in der Händigkeit der Untersuchenden liegen. Jedoch besteht in konkretem Fall bei beiden dieselbe Händigkeit (rechts), wodurch diese Annahme zu verwerfen ist.

Möglicherweise widerspiegelt das Ergebnis lediglich eine Präferenz von T2 zur rechten Seite.

#### **7.2.5 Interpretation des Zusammenhangs zwischen Ergebnissen und Händigkeit**

Da dem Flexionstest in der bisherigen Literatur nur wenig Beachtung zugekommen ist, findet sich auch kein Hinweis auf einen möglichen Zusammenhang zwischen einem positiven Testergebnis einer Rippe und der Händigkeit der Probandin bzw. des Probanden. Die Ergebnisse dieser Studie lassen ebenso keine eindeutige Aussage hinsichtlich eines Zusammenhangs zu, da die tatsächliche Übereinstimmung für eine auffällige Seite in nur sieben Fällen gegeben war. Diese Übereinstimmungen könnten auch per Zufall entstanden sein.

### **7.3 Bedeutung der Ergebnisse im Kontext der Osteopathie**

Aufgrund des niedrigen Kappa Koeffizienten von  $\kappa = 0.04$  sollte die Anwendung des Flexionstests im klinischen Alltag kritisch reflektiert werden. Obwohl sich der Test durch eine rasche Durchführbarkeit in einfacher Ausgangsposition in der osteopathischen Praxis auszeichnet, lässt das Ergebnis der Studie keine Empfehlung für die Verwendung des Flexionstests zu. Die erhobenen Daten sprechen für den Schluss, den Test nicht anzuwenden.

### **7.4 Kritische Reflexion der angewandten Methodik**

Die erhobenen Cohens Kappa Koeffizienten für die gesamte Stichprobe (sowohl mit zwei als auch mit drei Ergebniskategorien) und die Koeffizienten unter Berücksichtigung des Geschlechts der Teilnehmenden deuten alle auf eine geringe Übereinstimmung der Testerin und des Testers hin.

In vorliegender Studie wurde Wert darauf gelegt, den Ansprüchen der internen als auch der externen Validität zu genügen. Dabei entstandene Limitationen in der angewandten Methodik sollen folglich kritisch beleuchtet werden, um mögliche Ursachen für diese niedrigen Werte zu eruieren. Ferner sollen auch positive Aspekte vorliegender Studie angeführt werden.

Zur besseren Übersicht werden die Einflüsse auf die Literaturrecherche getrennt von den Einflüssen auf das Studiendesign, die Stichprobe sowie das Studienpersonal angeführt.

#### **7.4.1 Kritische Reflexion bezüglich Literaturrecherche**

Die in Kapitel 5.1 beschriebene Literaturrecherche wurde von Seiten der Studienautorin nach bestem Wissen und Gewissen durchgeführt, unterlag möglicherweise dennoch folgenden potenziellen Verzerrungen:

Die Suche nach Literatur wurde auf deutsch- und englischsprachige Publikationen beschränkt, wodurch laut Scherfer et al. (2011, S. 138) ein Language Bias entstehen kann. Relevante anderssprachige Literatur wurde demnach in der Recherche vernachlässigt, was Lücken in der Darstellung des aktuellen Forschungsstandes mit sich gebracht haben könnte.

Die Literaturrecherche umfasste zudem ausschließlich bereits publizierte Studien. Vor allem Studien mit statistisch nicht signifikanten Ergebnissen bleiben oftmals unveröffentlicht (Döring und Bortz, 2016, S. 901) oder werden ausschließlich in grauer Literatur publiziert (Hussy et al., 2013, S. 161). Diese Studien wurden somit in der Recherche nicht erfasst. Dadurch könnte ein verzerrter Eindruck der Ergebnisse entstanden sein, was als Publication Bias bezeichnet wird (Döring und Bortz, 2016, S. 901).

Wie in Kapitel 5.1 bereits erwähnt, konnte ein möglicher Retrieval Bias (Scherfer et al., 2011, S. 60) durch eine umfassende Literaturrecherche mit der Verwendung zahlreicher Suchbegriffe vermieden werden.

#### **7.4.2 Kritische Reflexion bezüglich Studiendesign**

Um die Anforderungen an die interne Validität zu einem hohen Grad erfüllen zu können, wurde der Studienablauf streng standardisiert (siehe Kapitel 5). Trotz dem hohen Maß an Standardisierung waren potenzielle, die Qualität vorliegender Studie betreffende, Limitationen nicht vermeidbar:

Das Untersuchen von mehreren Testverfahren im Zuge einer einzigen Reliabilitätsstudie könnte sich laut Patijn (2004) als nachteilig herausstellen. Dies gilt jedoch nur, wenn mehrere Tests durchgeführt werden, welche das Vorhandensein ein und derselben Diagnose beurteilen (Patijn, 2019). Laut Autor könnte dies in weiterer Folge dazu führen, dass die Testerin bzw. der Tester nach einem positiven Test A unbewusst ebenso den Test B als positiv bewertet und daraus eine gegenseitige Abhängigkeit der beiden Testverfahren entsteht. Somit wird hierbei nicht die Reliabilität jedes einzelnen Testverfahrens geprüft, sondern vielmehr die Reliabilität der Kombination dieser Tests (Patijn, 2004).

In vorliegender Studie wurden zwar zwei manuelle Tests in einem ähnlichen Areal durchgeführt, jedoch zielen beide Untersuchungen auf unterschiedliche funktionelle Dysfunktionen ab. Eine blockierte erste Rippe kann ein Thoracic Outlet Syndrome hervorrufen (Corts, 2019, S. 32), dies muss aber nicht zwangsläufig so sein. Ebenso gibt es Patientinnen und Patienten mit TOS, bei denen keine Dysfunktion der ersten Rippe vorliegt (Hooper et al., 2010).

Nichtsdestotrotz könnte in vorliegender Studie eine gegenseitige Beeinflussung der Tests vorgelegen sein. Diese mögliche Verzerrung der Ergebnisse wird als Confirmation Bias bezeichnet. Lieb et al. (2011, S. 30) beschreiben diese Art von Verzerrung als „erhöhte Bereitschaft, Informationen zu beachten, die zu einer bereits gefassten Erwartung passen“.

Um den oben beschriebenen Einfluss von einem Test auf den anderen möglichst gering zu halten, wurde in vorliegender Studie auf eine Randomisierung der Teilnehmerinnen und Teilnehmer sowie auf die Anwendung von randomisierten Testreihenfolgen (siehe Kapitel 5.9) gesetzt. Möglicherweise hätte die Studie trotzdem andere Ergebnisse hervorgebracht, wenn nur ein manueller Test auf seine Verlässlichkeit hin überprüft worden wäre.

Die Problematik des Auffindens des Processus spinosus von Th1 sowie der ersten Rippe wurde bereits in Kapitel 7.2.1 thematisiert. An dieser Stelle soll die gemeinhin schlecht belegte Reliabilität bezüglich der Palpation von ossären Strukturen an der Wirbelsäule aufgegriffen werden. In einer Übersichtsarbeit von Haneline und Young (2009) wurden unter anderem Reliabilitätsstudien bezüglich der Palpation von ossären Orientierungspunkten an der Wirbelsäule analysiert. Das Ergebnis zeigte vorwiegend niedrige Werte für die Interrater sowie Intrarater Reliabilität.

Dennoch wurde in vorliegender Studie die Markierung des Processus spinosus von Th1 durchgeführt, um sicherstellen zu können, dass zumindest beide Untersuchungspersonen bei der Palpation der ersten Rippe von derselben Struktur ausgehen.

Die Problematik hinsichtlich der Beurteilungskriterien sowie der strikten Vorgabe, nur einmal testen zu dürfen, wurde bereits in Kapitel 7.2.1 angeführt. Diese Vorgabe für die limitierte Anzahl der Untersuchungsgänge könnte als praxisfern bezeichnet werden, denn im klinischen Alltag kann die Testung uneingeschränkt oft wiederholt werden. Die strenge Limitation wurde dennoch gewählt, um dem Anspruch der Standardisierung zu genügen. Außerdem kann, wie in Kapitel 2.2.1 erwähnt, ein mehrmaliges Testen zu einem mobilisierenden Effekt führen, was folglich einen Einfluss auf die Testung durch die zweite Untersuchungsperson gehabt hätte.

Damit die Ergebnisse der Studie trotz des hohen Maßes an Standardisierung dennoch auf den klinischen Alltag übertragen werden können, wurde versucht, den Testablauf möglichst praxisnahe zu gestalten. Dazu wurde unter anderem von einer tatsächlichen optischen Verblindung (z.B. durch verbundene Augen der Testerin bzw. des Testers) abgesehen. Denn solch eine Verblindung stellt sich bei dieser Art von Test als praxisfern dar, zumal die Inspektion auch im osteopathischen Untersuchungsgang vor dem Testen der ersten Rippe standardmäßig durchgeführt wird. Eine optische Verblindung der Testerin bzw. des Testers hätte somit die Übertragbarkeit der Ergebnisse in den klinischen Alltag herabgesetzt, weshalb in vorliegender Studie davon abgesehen wurde.

#### **7.4.3 Kritische Reflexion bezüglich Stichprobe**

Die Größe der Stichprobe wurde aufgrund der Empfehlung von Patijn (2019) auf 40 Probandinnen und Probanden festgelegt. Eine weitere Möglichkeit wäre laut dem Autor gewesen, eine genaue Fallzahlberechnung zu machen.

Die 95% KI der vier Kappa Werte sind weit gefasst, was aus der Größe der Stichprobe ( $n = 40$ ) resultiert. Je kleiner die Stichprobe ist, desto weiter ist das dazugehörige Konfidenzintervall gefasst (Scherfer et al., 2011, S. 209). Möglicherweise hätte eine Fallzahlberechnung die optimale Stichprobengröße mit mehr als 40 Probandinnen und Probanden hervorgebracht, wodurch sich die Spanne der Konfidenzintervalle verkleinert hätte. Dadurch hätte eventuell die Aussagekraft der Ergebnisse erhöht werden können.

Wie in Kapitel 5.3.5 beschrieben, war die verwendete Stichprobe hinsichtlich ihrer Geschlechterverteilung nicht homogen (28 Frauen, 12 Männer). Ebenso gab es eine Diskrepanz in der Anzahl der symptomatischen und asymptomatischen Probandinnen und Probanden (13 symptomatisch, 27 asymptomatisch).

Jedoch wurde bewusst keine perfekte Homogenität der Stichprobe angestrebt, da solch eine Einheitlichkeit auch in der klinischen Praxis nicht realistisch ist und dadurch die externe Validität der Studie herabgesetzt worden wäre.

#### **7.4.4 Kritische Reflexion bezüglich Studienpersonal**

Die für die Studie gewählten Untersuchungspersonen verfügten zum Zeitpunkt der Testung über denselben Ausbildungsstand und dieselbe Berufserfahrung, was die Vergleichbarkeit zwischen den beiden deutlich erhöht. Zudem wurde durch die Verblindung zwischen der Testerin und dem Tester das Risiko einer möglichen Verzerrung der Ergebnisse gesenkt (siehe Kapitel 5.8).

Das hohe Maß an Standardisierung erforderte unter anderem das Durchlaufen einer Trainingsphase der Testerin und des Testers vor der eigentlichen Testung. Diese intensive Auseinandersetzung mit dem Flexionstest entspricht nicht dem Alltag der Osteopathinnen und Osteopathen, welche den Test in der Ausbildung erlernen und gleich anschließend (ohne Trainingsphase) in der Praxis umsetzen. Laut Van Trijffel et al. (2005) wird auch in der Literatur ein Zusammenhang zwischen der Trainingsphase und einer verminderten externen Validität von Studien diskutiert. Die Autorinnen und Autoren konnten diese These in ihrem systematischen Review jedoch nicht bestätigen. Daher bleibt auch die Frage offen, ob in vorliegender Studie solch ein potenzieller Einfluss auf die externe Validität besteht.

## 8 Konklusion

Das Kapitel soll neben einer kurzen Zusammenfassung der vorliegenden Studie auch einen Ausblick für weitere, diesen Untersuchungsgegenstand betreffende, Studien liefern.

Der Flexionstest stellt ein gängiges Untersuchungsverfahren für die Mobilität der ersten Rippe dar und zeichnet sich durch die rasche Durchführbarkeit in einfacher Ausgangsposition aus. In der bisherigen Forschung wurde die Testung weder auf ihre Reliabilität noch auf ihre Validität hin untersucht. Ziel dieser Arbeit war es daher, Werte für die Interrater Reliabilität des Flexionstests zu erheben, um herauszufinden, wie hoch die Übereinstimmung zweier unabhängiger Untersucherinnen bzw. Untersucher ist.

Im Zuge der Studie wurden 40 symptomatische und asymptotische Probandinnen und Probanden von zwei unabhängigen Untersuchungspersonen getestet. An derselben Stichprobe wurde im Zuge einer parallel durchgeführten Studie das costoclaviculäre Manöver auf seine Verlässlichkeit hin untersucht.

Um eine gegenseitige Beeinflussung der beiden Testverfahren auszuschließen, wurden die beiden Tests in unterschiedlichen randomisierten Reihenfolgen an den Probandinnen und Probanden appliziert. Während des gesamten Studienablaufes wurde eine Verblindung zwischen der Testerin und dem Tester sowie eine Verblindung zwischen den Untersuchenden und den Getesteten garantiert.

Die Studienleitung markierte an den Studienteilnehmerinnen und -teilnehmern den Processus spinosus des Wirbels Th1. Anschließend wurde der Flexionstest von jeder der beiden Untersuchungspersonen durchgeführt und mit einer der drei Kategorien (unauffällig, links auffällig, rechts auffällig) bewertet.

Aus den erhobenen Daten wurde mittels Cohens Kappa Koeffizienten der Grad der Übereinstimmung berechnet. Der Wert für alle Teilnehmerinnen und Teilnehmer liegt bei  $\kappa = 0.04$ , für die Probandinnen bei  $\kappa = 0.03$  und für die Probanden bei  $\kappa = 0.06$ . Alle Werte deuten auf eine geringfügige Übereinstimmung der Testerin und des Testers hin (Landis & Koch, 1977, S. 165).

Somit konnte die Forschungsfrage – Wie hoch ist die Übereinstimmung der Ergebnisse des Flexionstests bei symptomatischen sowie asymptotischen Erwachsenen durchgeführt von zwei unabhängigen Untersuchungspersonen? – wie folgt beantwortet werden: Die Übereinstimmung ist mit  $\kappa = 0.04$  sehr niedrig, was laut Landis und Koch (1977, S. 165) einer geringfügigen Übereinstimmung der Untersuchenden entspricht.

Aufgrund der Ergebnisse kann die Nullhypothese  $H_0$  – Die Übereinstimmung der Ergebnisse des Flexionstests für die erste Rippe ist nicht substantiell (Landis & Koch, 1977, S. 165), der Cohens Kappa Koeffizient  $\kappa$  ist somit kleiner als 0.61 – angenommen werden.

Die Alternativhypothese  $H_1$  – Die Übereinstimmung der Ergebnisse des Flexionstests für die erste Rippe ist mindestens substantiell (Landis & Koch, 1977, S. 165), der Cohens Kappa Koeffizient  $\kappa$  ist somit größer bzw. gleich 0.61 – hingegen wird verworfen.

Aus der Diskussion in Kapitel 7 lassen sich folgende Empfehlungen für mögliche Folgestudien ableiten:

Die in vorliegender Studie angewandte Methodik sollte größtenteils beibehalten werden. Vor allem das Vorhandensein einer unabhängigen Studienleitung, die der tatsächlichen Testung vorangehende Trainingsphase, der standardisierte Ablauf der Testung, die Verblindung zwischen der Testerin bzw. dem Tester, die Verblindung zwischen den Teilnehmenden und den Untersuchenden sowie die Randomisierung der Testreihenfolge haben sich bewährt und sollten daher in zukünftigen Studien beibehalten werden.

Möglicherweise sollte eine Reliabilitätsstudie, welche sich ausschließlich einem einzigen Test – nämlich dem Flexionstest – widmet, durchgeführt werden.

Zudem sollte bei zukünftigen Studien über eine Fallzahlberechnung die genaue Größe der Stichprobe ermittelt werden.

Neben der Markierung des Processus spinosus von Th1 würde sich potenziell auch eine Markierung der ersten Rippe als erleichternd für die Untersuchenden erweisen.

Weiters könnte der BMI der Probandinnen und Probanden erhoben werden, um einen etwaigen Einfluss auf die Studienergebnisse sichtbar machen zu können.

Um eine finale Beurteilung des Flexionstests abgeben zu können, wäre es neben der erhobenen Interrater Reliabilität essenziell, die Intrarater Reliabilität des Testverfahrens ebenfalls zu untersuchen. Zusätzlich könnte der Flexionstest für die Mobilität der ersten Rippe auch auf das Gütekriterium der Validität hin überprüft werden. Jedoch wird von Messinstrumenten, welche den hohen Anforderung der Reliabilität nicht entsprechen, ebenso keine gute Validität erwartet (Scherfer et al., 2011, S. 244).

Zusammenfassend kann keine Empfehlung für die Anwendung des Flexionstests im klinischen Alltag gegeben werden.

## LITERATURVERZEICHNIS

- Bortz, J., & Lienert, G. A. (2008). *Kurzgefasste Statistik für die klinische Forschung: Leitfaden für die verteilungsfreie Analyse kleiner Stichproben; mit 97 Tabellen sowie zahlreichen Formeln* (3., aktualisierte und bearb. Aufl.). Springer Medizin.
- Breidenbach, O., & Erlewein, R. (2015). *Lehrbuch Dorn-Therapie: Mit Skribben und Breuß-Massage* (4., überarb. und erw. Aufl.). Karl F. Haug.
- Corts, M. (2016). *Diagnoseleitfaden Osteopathie* (2. überarbeitete und erweiterte Edition). Karl F. Haug.
- Corts, M. (2019). *Anatomie für Osteopathen: Lehrbuch und Atlas*. Georg Thieme Verlag.
- Diemer, F., & Sutor, V. (2011). *Praxis der medizinische Trainingstherapie I: Lendenwirbelsäule, Sakroiliakgelenk und untere Extremität* (2. Edition). Thieme.
- Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Springer-Verlag.
- Fjellner, A., Bexander, C., Faleij, R., & Strender, L.-E. (1999). Interexaminer reliability in physical examination of the cervical spine. *Journal of Manipulative and Physiological Therapeutics*, 22(8), 511–516. [https://doi.org/10.1016/S0161-4754\(99\)70002-3](https://doi.org/10.1016/S0161-4754(99)70002-3)
- Greenman, P. E. (2005). *Lehrbuch der Osteopathischen Medizin* (3., überarbeit. und erweiter. A. Edition). Karl F. Haug.
- Haneline, M. T., & Young, M. (2009). A Review of Intraexaminer and Interexaminer Reliability of Static Spinal Palpation: A Literature Synthesis. *Journal of Manipulative and Physiological Therapeutics*, 32(5), 379–386. <https://doi.org/10.1016/j.jmpt.2009.04.010>
- Heiderscheit, B., & Boissonnault, W. (2008). Reliability of Joint Mobility and Pain Assessment of the Thoracic Spine and Rib Cage in Asymptomatic Individuals. *Journal of Manual & Manipulative Therapy*, 16(4), 210–216. <https://doi.org/10.1179/106698108790818369>

- Hollerwöger, D. (2006). Methodological quality and outcomes of studies addressing manual cervical spine examinations: A review. *Manual Therapy*, 11(2), 93–98. <https://doi.org/10.1016/j.math.2005.12.001>
- Hooper, T. L., Denton, J., McGalliard, M. K., Brismée, J.-M., & Sizer, P. S. (2010). Thoracic outlet syndrome: A controversial clinical condition. Part 1: anatomy, and clinical examination/diagnosis. *Journal of Manual & Manipulative Therapy*, 18(2), 74–83. <https://doi.org/10.1179/106698110X12640740712734>
- Hussy, W., Schreier, M., & Echterhoff, G. (2013). *Forschungsmethoden in Psychologie und Sozialwissenschaften für Bachelor*. Springer-Verlag.
- Hüter-Becker, A., & Dölken, M. (2015). *Physiotherapie in der Orthopädie* (3. Edition). Thieme.
- Krumholz, L. (1998). Technik zur Untersuchung der ersten Rippe. *Manuelle Medizin*, 36(5), 246–251. <https://doi.org/10.1007/s003370050082>
- Laekeman, M., & Kreutzer, R. (2009). *Großer Bildatlas der Palpation: Anatomische Strukturen gezielt lokalisieren und begreifen*. Springer-Verlag.
- Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>
- Langer, W., & Hebggen, E. (2012). *Lehrbuch Osteopathie* (1. Edition). Haug Fachbuch.
- Lieb, K., Klemperer, D., & Ludwig, W.-D. (2011). *Interessenkonflikte in der Medizin: Hintergründe und Lösungsmöglichkeiten*. Springer-Verlag.
- Liem, T., & Dobler, T. K. (2002). *Leitfaden Osteopathie*. Urban & Fischer Bei Elsevier.
- Lindgren, K.-A., Leino, E., & Manninen, H. (1992). Cervical rotation lateral flexion test in brachialgia. *Archives of Physical Medicine and Rehabilitation*, 73(8), 735–737. <https://doi.org/10.5555/uri:pii:000399939290208E>
- Mastromarchi, P., & May, S. (2020). First rib dysfunction in patients with neck and shoulder pain: A Delphi investigation. *Journal of Manual & Manipulative Therapy*, 1–8. <https://doi.org/10.1080/10669817.2020.1824470>

- Matzen, P. A., Alzen, G., & Matzen, P. (Hrsg.). (2007). *Kinderorthopädie: Mit 73 Tabellen; [mit 1150 diagnostischen Abbildungen]* (1. Aufl). Elsevier, Urban & Fischer.
- Michael, A. (1995). First Rib Function and the Thoracic Outlet Syndrome. *Journal of Orthopaedic Medicine*, 17(2), 56–61.  
<https://doi.org/10.1080/1355297X.1995.11719787>
- Mitchell, F. L., & Mitchell, P. K. G. (2004). *Handbuch der MuskelEnergieTechniken: Diagnostik und Therapie: Brustkorb, Brust- und Lendenwirbelsäule / übers. von Markus Vieten. Bd. 2.* Georg Thieme Verlag.
- Nomden, J. G., Slagers, A. J., Bergman, G. J. D., Winters, J. C., Kropmans, T. J. B., & Dijkstra, P. U. (2009). Interobserver reliability of physical examination of shoulder girdle. *Manual Therapy*, 14(2), 152–159. <https://doi.org/10.1016/j.math.2008.01.005>
- Patijn, J. (Hrsg.). (2004). Reproducibility and validity studies of diagnostic procedures in manual/musculoskeletal medicine. *Musculoskeletal Medicine*, 3, 2–35.
- Patijn, J. (2019). Reproducibility protocol for diagnostic procedures in Manual/Musculoskeletal Medicine: Edition 2019. *Manuelle Medizin*, 57(6), 451–479.  
<https://doi.org/10.1007/s00337-019-00581-5>
- Peeters, L., & Lason, G. (2015). *Osteopathische Medizin- Die Brustwirbelsäule und die Rippen.* Osteo 2000 bvba.
- Póvoa, L. C., Ferreira, A. P. A., Zanier, J. F. C., & Silva, J. G. (2018). Accuracy of Motion Palpation Flexion-Extension Test in Identifying the Seventh Cervical Spinal Process. *Journal of Chiropractic Medicine*, 17(1), 22–29.  
<https://doi.org/10.1016/j.jcm.2017.11.005>
- Reichert, B. (Hrsg.). (2018). *Palpations-Techniken: Anatomie in vivo.* Georg Thieme Verlag.
- Rieder, A., & Lohff, B. (Hrsg.). (2008). *Gender Medizin: Geschlechtsspezifische Aspekte für die klinische Praxis* (2., überarb. und erw. Aufl). Springer-Verlag.

- Robinson, R., Robinson, H. S., Bjørke, G., & Kvale, A. (2009). Reliability and validity of a palpation technique for identifying the spinous processes of C7 and L5. *Manual Therapy*, 14(4), 409–414. <https://doi.org/10.1016/j.math.2008.06.002>
- Schäfer, A., & Schöttker-Königer, T. (2015). *Statistik und quantitative Methoden für Gesundheitsfachberufe*. Springer-Verlag.
- Scherfer, E., Bossmann, T., & Herbert, R. D. (2011). *Forschung verstehen: Ein Grundkurs in evidenzbasierter Praxis* (2., überarbeitete und erweiterte Auflage). Pflaum Verlag.
- Smedmark, V., Wallin, M., & Arvidsson, I. (2000). Inter-examiner reliability in assessing passive intervertebral motion of the cervical spine. *Manual Therapy*, 5(2), 97–101. <https://doi.org/10.1054/math.2000.0234>
- Streeck, U., Focke, J., Klimpel, L. D., & Noack, D.-W. (2006). *Manuelle Therapie und komplexe Rehabilitation: Band 1: Grundlagen, obere Körperregionen*. Springer-Verlag.
- Strunk, A. (2012). *Fasziale Osteopathie: Grundlagen und Techniken* (1. Edition). Karl F. Haug.
- van Trijffel, E., Anderegg, Q., Bossuyt, P. M. M., & Lucas, C. (2005). Inter-examiner reliability of passive assessment of intervertebral motion in the cervical and lumbar spine: A systematic review. *Manual Therapy*, 10(4), 256–269. <https://doi.org/10.1016/j.math.2005.04.008>
- Vaught, M. S., Brismée, J.-M., Dedrick, G. S., Sizer, P. S., & Sawyer, S. F. (2011). Association of Disturbances in the Thoracic Outlet in Subjects with Carpal Tunnel Syndrome: A Case–Control Study. *Journal of Hand Therapy*, 24(1), 44–52. <https://doi.org/10.1016/j.jht.2010.09.070>
- Ward, R. C. (Hrsg.). (2003). *Foundations for osteopathic medicine* (2nd ed). Lippincott Williams & Wilkins.

## TABELLENVERZEICHNIS

<i>Nummer</i>	<i>Seite</i>
<i>Tabelle 1. Kappa Koeffizienten für Mobilitätstests der ersten Rippe in der Literatur .....</i>	<i>14</i>
<i>Tabelle 2. Kreuztabelle .....</i>	<i>28</i>
<i>Tabelle 3. Klassifikation der Cohens Kappa Koeffizienten .....</i>	<i>29</i>
<i>Tabelle 4. Kreuztabelle – alle Probandinnen und Probanden – drei Ergebniskategorien .....</i>	<i>30</i>
<i>Tabelle 5. Kreuztabelle – alle Probandinnen und Probanden – zwei Ergebniskategorien .....</i>	<i>31</i>
<i>Tabelle 6. Kreuztabelle – Probandinnen – drei Ergebniskategorien .....</i>	<i>31</i>
<i>Tabelle 7. Kreuztabelle – Probanden – drei Ergebniskategorien .....</i>	<i>32</i>
<i>Tabelle 8. Übersicht über alle errechneten Cohens Kappa Koeffizienten mit zugehörigem 95% KI .....</i>	<i>32</i>
<i>Tabelle 9. Häufigkeitsverteilung der Testergebnisse .....</i>	<i>33</i>
<i>Tabelle 10. Geschlechtsverteilung der auffälligen Testergebnisse .....</i>	<i>33</i>

## ABBILDUNGSVERZEICHNIS

<i>Nummer</i>	<i>Seite</i>
<i>Abbildung 1. Anteil der symptomatischen und Studienteilnehmerinnen – und teilnehmer .....</i>	<i>21</i>
<i>Abbildung 2. Endposition des Flexionstests .....</i>	<i>26</i>

## ABKÜRZUNGSVERZEICHNIS

AP	anterior- posterior gerichteter Druck
BMI-Wert	Body Mass Index Wert
CRLF-Test	Cervical Rotation Lateral Flexion Test
PA	posterior-anterior gerichteter Druck
T1/T2	Testpersonal
TOS	Thoracic Outlet Syndrome
95% KI	95% Konfidenzintervall

## ANHANG A

### Studieninformation

**Thema:** Interrater Reliabilität zur Testung der Mobilität der ersten Rippe und des Costoclaviculären Manövers

**Studienautorin/Studienautor:** Kondler Verena, Puhr Maximilian

**Zeitpunkt:** Samstag, 24.4.2021

**Ort:** Praxis Osteopathie Gröbming – 1. Stock, Hauptplatz 110, 8962 Gröbming

**Kurzbeschreibung:** Im Zuge unserer Masterarbeiten untersuchen wir die Verlässlichkeit zweier manualtherapeutischer Tests. Konkret geht es um die Frage, ob zwei Untersuchende bei der Durchführung desselben Tests zum gleichen Ergebnis kommen.

Bei der ersten Untersuchung – Testung der Mobilität der ersten Rippe – geht es darum, mögliche Immobilitäten der Rippe im Seitenvergleich festzustellen.

Die zweite Testung – das Costoclaviculäre Manöver – untersucht eine mögliche Einengung des Nervengeflechts im Bereich der oberen Extremität im Seitenvergleich.

Beide Untersuchungen stellen gängige Diagnose-Verfahren in der Physiotherapie und in der Osteopathie dar.

Bei der Studie handelt es sich um eine reine Untersuchung und keine Behandlung bzw. therapeutische Intervention.

**Organisatorisches:** Beide Untersuchungen werden ausschließlich mittels Palpation (Spüren) durchgeführt. Es ist notwendig, dass Sie Ihren Oberkörper bis auf die Unterwäsche frei machen, um die betreffende Region sicher auffinden und beurteilen zu können.

Vor Beginn der eigentlichen Testung wird der erste Brustwirbel mit einem Hautstift markiert.

Anschließend werden die zwei Tests von beiden Untersuchenden durchgeführt.

Um die Studienergebnisse nicht zu verfälschen, sollten die Gespräche während der Testung auf ein Minimum reduziert werden. Wir bitten um Ihr Verständnis.

**Nebenwirkungen:** Bei der Testung sind keinerlei Nebenwirkungen zu erwarten. Sollten Sie innerhalb von 24 Stunden nach Durchführung der Testung Unwohlsein verspüren, melden Sie sich bitte telefonisch bei uns unter der Nummer [REDACTED].

**Datenschutz:** Alle erhobenen Daten werden selbstverständlich ausschließlich in anonymisierter Form ausgewertet und weiterverarbeitet.

Wenn Sie an Ihren individuellen Testergebnissen interessiert sind, geben wir Ihnen gerne im Laufe der nachfolgenden Tage darüber Auskunft.

**Hygienemaßnahmen/COVID-Richtlinien:** Aufgrund der aktuellen COVID-Situation bitten wir Sie, die Testräumlichkeiten ausschließlich mit einer FFP2-Maske zu betreten. Um größere Menschenansammlungen zu vermeiden, soll der Ihnen zugewiesene Termin bitte pünktlich eingehalten werden. Dieser Termin wird zeitnah bekanntgegeben.

**Sonstiges:** Bei Fragen oder Unklarheiten zum Studienablauf melden Sie sich bitte per Mail an [REDACTED] oder [REDACTED].

Wir bitten Sie den Fragebogen bzw. die Einverständniserklärung ausgefüllt in Papierform zur Testung mitzubringen.

Wir bedanken uns herzlich bei Ihnen für Ihre Bereitschaft zur Teilnahme an der Studie.

## ANHANG B

### Fragebogen bezüglich Ausschlusskriterien und Einverständniserklärung

Name: .....

weiblich  männlich  Alter: .....

Hatten Sie in den vergangenen 6 Wochen **Verletzungen** in folgenden Bereichen:  
Halswirbelsäule, Brustwirbelsäule, Rippen, Schlüsselbeine, obere Extremitäten?

ja  nein

Wenn ja, welche? .....

Besteht eine **ärztlich vorgeschriebene Bewegungseinschränkung** in folgenden Bereichen:  
Halswirbelsäule, Brustwirbelsäule, Rippen, Schlüsselbeine, obere Extremitäten?

ja  nein

Wenn ja, welche? .....

Leiden Sie unter einem diagnostizierten **Aneurysma**?

ja  nein

Leiden Sie unter **Herzinsuffizienz** oder anderen **Herzerkrankungen**?

ja  nein

Wenn ja, welche? .....

Falls Sie eine der Fragen mit JA beantwortet haben, bitten wir Sie um kurze Kontaktaufnahme mit uns.

Mit meiner Unterschrift bestätige ich Folgendes:

- die Richtigkeit meiner angegebenen Daten
- die freiwillige Teilnahme an der Studie
- ausreichend über Inhalt, Ablauf, Risiken und Ziel der Studie aufgeklärt worden zu sein
- die Weiterverwendung meiner Daten in anonymisierter Form zur Auswertung
- die Möglichkeit, zu jedem Zeitpunkt von der Teilnahme an der Studie zurückzutreten

Ort, Datum: .....

Unterschrift der Teilnehmerin/des Teilnehmers: .....

## ANHANG C

### Fragebogen bezüglich Händigkeit und Symptomatik

Proband\*in-Nummer:

**Ich bin**      linkshändig       rechtshändig       beidhändig

Bei mir treten fallweise folgende Symptome im Schultergürtel, den Armen und/oder Händen auf:

#### **Schmerzen**

nein       ja, rechts       ja, links

Wenn ja, verstärken sich diese Schmerzen **bei tiefer Einatmung bzw. Ausatmung?**

nein       ja

Wenn ja, verstärken sich diese Schmerzen bei Tätigkeiten wie **Heben oder Tragen?**

nein       ja

#### **Schwellung**

nein       ja, rechts       ja, links

#### **Kribbeln / einschlafendes Gefühl**

nein       ja, rechts       ja, links

#### **Kraftlosigkeit**

nein       ja, rechts       ja, links

## ANHANG D

### **Test-Instruktion**

Bevor wir starten, ist es wichtig, dass Sie jederzeit Bescheid geben, wenn Ihnen etwas unangenehm ist oder Schmerzen bereitet.

Melden Sie sich – dann wird die Testung sofort gestoppt.

Bitte setzen Sie sich entspannt hin und lehnen Sie sich an der Rückenlehne an.

Ich werde nun auf die erste Rippe greifen und auf mein Kommando bewegen Sie ihr Kinn so weit wie möglich zum Brustkorb und rollen den Kopf ein.

# ANHANG E

## Beurteilungsbogen

Proband\*in-Nummer:

<b>Flexionstest 1. Rippe</b>	
links auffällig	rechts auffällig

ANHANG F

Rohdaten der Testung am 24.04.2021

Proband/Probandin	Geschlecht	Testerin/Tester 1	Testerin/Tester 2	Proband/Probandin	Geschlecht	Testerin/Tester 1	Testerin/Tester 2
1	♀	1	2	21	♀	1	1
2	♀	1	1	22	♀	1	2
3	♀	2	2	23	♀	0	2
4	♂	1	2	24	♂	1	0
5	♀	1	2	25	♂	0	0
6	♂	0	2	26	♀	2	2
7	♀	1	2	27	♀	2	0
8	♀	1	0	28	♀	2	0
9	♀	2	0	29	♂	1	2
10	♂	2	0	30	♀	1	2
11	♀	2	2	31	♂	1	2
12	♀	2	1	32	♀	1	0
13	♀	1	2	33	♀	0	0
14	♂	2	0	34	♀	1	0
15	♀	0	2	35	♂	0	0
16	♂	1	2	36	♀	0	0
17	♀	1	0	37	♀	0	2
18	♀	0	2	38	♂	0	0
19	♂	2	2	39	♀	0	2
20	♀	2	2	40	♀	0	0

## ANHANG G

### Englische Kurzfassung

#### **Interrater reliability of the flexion test for the first rib in symptomatic and asymptomatic adults**

Verena Kondler, BSc

Mag. Dr. Astrid Grant Hay

Danube University Krems – Department for Health Sciences and Biomedicine

#### **Correspondence Address**

**Name:** Verena Kondler, BSc

**Address:** Hofmanning 420/18, 8962 Gröbming

**E-Mail address:** veri.kondler@gmx.at

**Phone number:** 0664/3666290

#### **Curriculum Vitae**

The author Verena Kondler, BSc, attended the education for physiotherapy at the University of Applied Science Upper Austria in Linz from 2013 until 2016. She started to work as a physiotherapist at the rehabilitation centre Gröbming in Styria in summer 2016 and is currently still employed there. Additionally, the author is working in a private practice for physiotherapy and osteopathy in Gröbming since autumn 2018.

The author started the education for osteopathy in 2016 at the Vienna School of Osteopathy WSO. Currently, she is attending the master programme at the Danube University Krems.

## **ABSTRACT**

**Objective:** The flexion test is a common examination test for the mobility of the first rib. Previous research has not yet investigated the reliability or validity of that test procedure. The aim of the present study is to collect data about the interrater reliability of the flexion test, in order to find out how two independent testers agree.

**Methods:** The present study is an interrater reliability design. Two independent testers examined 40 symptomatic and asymptomatic adults with the flexion test.

**Results:** The data was calculated with the Cohen's Kappa coefficient. For all participants the coefficient is  $\kappa = 0.04$ . Just for female participants it is  $\kappa = 0.03$  and for male participants it is  $\kappa = 0.06$ . All the values indicate a slight agreement between the testers.

**Conclusion:** The low interrater reliability does not allow a recommendation for the use of the flexion test in clinical practice. For a final appraisal of the test it would be relevant to examine the quality criterion of intrarater reliability and validity as well.

**Keywords:** interrater reliability, flexion test, first rib, osteopathy

## INTRODUCTION

The first rib is responsible for a lot of different symptoms. Trauma, especially whiplash trauma, or carrying and lifting heavy weight unilaterally can lead to a dysfunction of the first rib [1]. That kind of dysfunction may cause pain in the posterior subscapular or anterior subclavicular region and might be dependent on respiration [2].

A dysfunction may also compress the plexus brachialis with the nerves and blood vessels of the upper extremity [3] resulting in emissive pain and paraesthesia in the upper extremity [4].

Depending on which structure gets compressed, it may lead to the following symptoms in the upper extremity [5]:

- paleness and the feeling of coldness (arterial compression)
- cyanosis, feeling of heaviness, oedema (venous compression)
- numbness, weakness, reduced fine motor skills (nerval compression)

Additionally there might be an involvement of the sympathetic chain [6] or the ganglion stellatum resulting in autonomic nervous reaction in the face or the upper extremity [3].

The plurality of different symptoms caused by a dysfunction of the first rib, requires an accurate and reliable motion test for the first rib, in order to find a possible dysfunction. The manual medicine offers a wide range of examination methods for the first rib. There are active and passive motion tests available.

Some of the passive motion tests of the first rib already got examined concerning their interrater reliability. These studies provide poor interrater reliability [7, 8, 9, 10]. Two of them showed a fair agreement [7, 8], two of them merely a slight agreement [9, 10]. On the contrary, the study which examined the CRLF test, showed a perfect agreement of the raters [11]. Though an insufficient number of participants took part in the study.

The previous research has not yet investigated reliability or validity of active motion tests of the first rib, e.g. the flexion test. The flexion test is characterised for its simple and practicable applicability. The bilateral testing makes it possible to examine both first ribs at the same time, which is time saving in clinical practice.

That research gap was taken as a cause for this study. The aim of the present study is to collect data about the interrater reliability of the flexion test, in order to find out how two independent testers agree.

## **METHODS**

### **Study design**

The present study is a methodological study, more precisely an interrater reliability study design. In the context of that study a second interrater reliability study with the same sample was done. While the present study examined the flexion test for the mobility of the first rib, the second study investigated the costoclavicular manoeuvre as a test for thoracic outlet syndrome.

### **Sample**

The study examined asymptomatic participants as well as participants with symptoms of a dysfunctional first rib. The minimum age was determined to be 18 years. The following criteria excluded from participating the study:

- acute bony injuries (< 6 weeks, due to incomplete healing) in the following parts of the body: cervical spine, clavícula, thoracic spine, ribs, upper extremity
- limitations of movement mandated by a doctor concerning the spine or the upper extremity, whereby the participant cannot achieve the starting or ending position of the test
- participants, who were examined and/or treated by one rater or even two of them in the past, in order to maintain the blinding of the study
- additional diagnosis like aneurysm or cardiac insufficiency, in order to not create an additive pressure on the cardiovascular system

The recruitment was carried out via an announcement targetting the staff of the rehabilitation centre Gröbming. Friends, acquaintances and relatives of the author were asked to participate as well.

In the present study 40 subjects took part (28 women, 12 men). All 40 participants finished the study. The average was  $40.1 \pm 13.8$  years (min = 18; max = 81).

13 people had symptoms of a dysfunctional first rib (32.5%), 27 people were asymptomatic (67.5%).

### **Raters**

Two raters were consulted. Both have been working as physiotherapists since 2016 and therefore have a similar amount of professional experience. Currently, both raters are attending the master programme for osteopathy at university.

### **Assistant**

The assistant of the study is working as a physiotherapist and osteopath as well. Her job was to ensure the blinding during the study process.

### **Blinding**

The blinding was assured by two different kinds of blinding: rater-rater blinding and participant-rater blinding.

The two raters were not allowed to talk to each other about the participants or their results during the test. Furthermore, they were not allowed to see into the list of results.

The second kind of blinding guaranteed that the raters did not know anything about the patient's symptoms. Therefore, the raters and participants were not allowed to talk to each other about potential symptoms or results of the test.

### **Training period**

Before the real phase of testing a training period was performed. The study staff reached an agreement about the standardised procedure concerning the following: starting position of the raters, starting position of the participants, anatomic landmarks as an alleviation for the palpation, handhold of the palpation hand, exact performance of testing, amount of revisions, verbal instruction of the participants and possible extend of interpretation in the rating of the test.

The training period was applied to 10 participants. After that, the raters were asked to apply the standardised procedure as often and on as many patients in daily practice as possible.

### **Examination**

The examination took place at the 24<sup>th</sup> of april 2021. Every 20 minutes two participants were scheduled. They were randomly assigned to one of the two raters.

The participants had to take off their shirt, except when the examined region was not covered with clothes in the first place. The assistant marked the vertebra Th 1 as an alleviation for the palpation. Subsequently the two tests (flexion test and costoclavicular manoeuvre) were applied one after another. Then the two raters changed the room and applied the tests once again. In order to avoid an impact from one test to other one, the order of application was chosen randomly.

The starting position for the flexion test was seat with the participant's back leant on the backrest. The rater was standing behind the participant and palpated with the thumbs the angulus costae of the first rib. Then the participant was asked to bend the head slowly and as far as possible. The thumbs were following the movement of the first rib to the cranial. If one

side was moving faster than the other, that side was rated as conspicuous. Three options for rating were available: right side conspicuous, left side conspicuous or both inconspicuous. Repeated testing was not allowed, as that might have had a mobilising effect on the rib and would therefore influence the test of the second rater.

## STATISTICS

The data was presented in a cross tabulation. The Cohen's Kappa coefficient was used to calculate the agreement between the two raters.

Beside a general Cohen's Kappa coefficient for the whole sample three more coefficients were calculated additionally:

- Cohen's Kappa coefficient for female participants
- Cohen's Kappa coefficient for male participants
- Cohen's Kappa coefficient for the whole sample with two categories (conspicuous or inconspicuous – without consideration of the sides)

For every coefficient the 95% confidence interval (= 95% CI) was also calculated. The coefficients were interpreted by the classification of Landis and Koch [12].

The distribution of elected categories as well as the distribution concerning gender were described in absolute numbers and in percentage.

## RESULTS

All calculated Cohen's Kappa coefficients are listed in Table 1. The lowest coefficient is the one for female participants ( $\kappa = 0.03$ ), the highest coefficient is the one for the whole sample with two categories ( $\kappa = 0.13$ ). All the values indicate a slight agreement between the testers [12].

**Table 1.** Summary of all Cohen's Kappa coefficients with 95% CI

<b>sample (n) – number of categories</b>	<b><math>\kappa</math></b>	<b>95% CI lower limit</b>	<b>95% CI upper limit</b>
whole sample (n = 40) – 3 categories	0.04	-0.15	0.23
whole sample (n = 40) – 2 categories	0.13	-0.18	0.44
female participants (n = 28) – 3 categories	0.03	-0.20	0.26
male participants (n = 12) – 3 categories	0.06	-0.24	0.36

Table 2 shows the distribution of elected categories of the two raters separately. It can be seen that the raters decided on the categories rather unequally.

**Table 2.** Distribution of elected categories

category	rater 1		rater 2	
	<i>absolute</i>	<i>percentage</i>	<i>absolute</i>	<i>percentage</i>
0	12	30%	16	40%
1	17	42.5%	3	7.5%
2	11	27.5%	21	52.5%

Table 3 shows the distribution concerning gender. If only the conspicuous findings are taken into account, both raters found more abnormalities in female participants than in male.

**Table 3.** Distribution concerning gender, only conspicuous findings

gender	rater 1		rater 2	
	<i>absolute</i>	<i>percentage</i>	<i>absolute</i>	<i>percentage</i>
♀	20	71%	18	64%
♂	8	67%	6	50%

## DISCUSSION

The highest value was the coefficient for the whole sample with two categories and without consideration of the sides ( $\kappa = 0.13$ ). However, the flexion test should not only evaluate if there is a conspicuity in any of the two ribs, it should rather identify the side of dysfunction. That is why that coefficient is clinically less relevant.

The calculated Cohen's Kappa coefficient for the whole sample ( $\kappa = 0.04$ ) is the lowest coefficient compared to the literature. Potential responsible causes will be listed below.

### Limiting factors concerning the study design

The present study was performed in cooperation with a second interrater reliability study. According to the literature an evaluation of more than one test could be disadvantageous, especially if both tests assess the same diagnosis [13]. In that case the result rather describes the reliability of the combination of tests than the reliability of every test itself [13].

This study examined two manual tests in a similar region; however, they were targeted on different dysfunctions. A blocked first rib may cause a thoracic outlet syndrome [3], yet must not necessarily. Equally, patients with thoracic outlet syndrome may not have a dysfunction in the first rib [5]. Nevertheless there might have been an unnoticed mutual influence from one

test to the other. In order to keep that influence as low as possible the tests were applied in randomised order.

The raters were not allowed to apply the motion test more than once, to avoid a mobilising effect, which could have had an impact on the test applied by the second rater [13]. Although the raters were not always a hundred percent certain after a single test, they had to decide on one category.

Despite the attempt to standardise, there was room for subjective interpretation concerning the results left. Sometimes it was difficult for the raters to decide on one of the three available categories, since the result of the test was not clear enough. This could also be an explanation for the low agreement of the raters.

For a good transfer of the results to clinical practice the present study did not use real visual blinding (e.g. blindfolded raters). A blinding would not be practical, as osteopaths perform an inspection by default before they apply a motion test of the first rib in practice.

Despite the standardised procedure there has been minimal differences in the starting position of the participants. In certain instances when the first rater applied the tests some of the participants were nervous and therefore had a hypertonus in the musculus trapezius pars descendens. When the second rater subsequently applied the tests, the participants already knew the procedure which led to a relaxed position. That circumstance may have created a slightly different starting position for the two raters.

The raters orientated themselves at the marked processus spinosus of vertebra Th1. The recent literature shows a low intrarater reliability as well as interrater reliability of the palpation of osseous structures of the spine [14]. Though the osseous landmark of vertebra Th1 was applied because of standardisation.

### **Limiting factors concerning the sample**

The size of the sample was determined to 40 participants by a recommendation from the literature [15]. There would have also been the option to calculate the exact number of cases.

The used sample was not homogenous concerning the distribution of gender. Moreover, there has been discrepancy regarding the number of symptomatic and asymptomatic participants. A perfect homogeneity of the sample was not even aspired to reach, as it is not realistic in clinical practice.

### **Limiting factors concerning the raters and the assistant**

Finding the processus spinosus of vertebra Th1 was not always clear for the assistant. As well as it was not always clear for the raters to find the angulus costae of the first rib. Especially in participants with a hypertonus of the musculus trapezius pars descendens or if augmented subcutaneous fat tissue covered the region, it was difficult to palpate the osseous structures. Therefore, there might have been differences in the performance of the test, which may be an explanation for the results of the study.

The subcutaneous fat tissue might also be a possible reason for the gender-specific differences in the results. The raters agreed more often in male participants than in female. This is due to women having a higher body fat percentage (20 to 30%) than men (10 to 20%) [16]. This coincides with the experience of the two raters, who reported that the palpation during the test was easier in male participants.

The reason for the difference concerning the distribution of elected categories is not clear, as both raters are right-handed. Maybe it merely reflects a preference from one rater to the right side.

### **CONCLUSION**

Because of the low Cohen's Kappa coefficients, the application of the flexion test in clinical practice should be reflected critically. Although the test is easy and fast to apply, the result of the present study does not allow a recommendation for the use of that test. The data moreover implies to not use the test in daily practice.

Following studies regarding the reliability of motion testing of the first rib should examine only one test. For a possible final appraisal of the flexion test one should examine the quality criterion of intrarater reliability and validity as well. In prospective studies an exact calculation of the number of cases is recommended in order to determine the perfect sample size. Furthermore, beside the landmark of processus spinosus of vertebra Th1 also a second landmark – the first rib – should be marked as an alleviation for the raters.

### **DISCLOSURE**

No conflicts of interest were reported and no funding was received for this study.

## REFERENCES

- [1] P. Mastromarchi and S. May, "First rib dysfunction in patients with neck and shoulder pain: a Delphi investigation," *Journal of Manual & Manipulative Therapy*, pp. 1-8, 2020.
- [2] M. Krumholz, "Technik zur Untersuchung der ersten Rippe," *Manuelle Medizin*, vol. 36, no. 5, pp. 246-251, 1998.
- [3] M. Corts, *Anatomie für Osteopathen: Lehrbuch und Atlas*, Stuttgart: Georg Thieme Verlag, 2019.
- [4] O. Breidenbach and R. Erlewein, *Lehrbuch Dorn-Therapie: Mit Skribben und Breuß-Massage*, Stuttgart: Karl F. Haug, 2015.
- [5] T. L. Hooper, J. Denton, P. S. Sizer, J.-M. Brismée and M. K. McGalliard, "Thoracic outlet syndrome: a controversial clinical condition. Part 1: anatomy, and clinical examination/diagnosis," *Journal of Manual & Manipulative Therapy*, vol. 18, no. 2, pp. 74-83, 2010.
- [6] A. Michael, "First Rib Function and the Thoracic Outlet Syndrome," *Journal of Orthopaedic Medicine*, vol. 17, no. 2, pp. 56-61, 1995.
- [7] J. G. Nomden, A. J. Slagers, G. J. Bergman, J. C. Winters, T. J. Kropmans and P. U. Dijkstra, "Interobserver reliability of physical examination of shoulder girdle," *Manual Therapy*, vol. 14, no. 2, pp. 152-159, 2009.
- [8] V. Smedmark, M. Wallin and I. Arvdisson, "Inter-examiner reliability in assessing passive intervertebral motion of the cervical spine," vol. 5, no. 2, pp. 97-101, 2000.
- [9] A. Fjellner, C. Bexander, R. Faleij and L.-E. Strender, "Interexaminer reliability in physical examination of the cervical spine," *Journal of Manipulative and Physiological Therapeutics*, vol. 22, no. 8, pp. 511-516, 1999.
- [10] B. Heiderscheit and W. Boissonnault, "Reliability of Joint Mobility and Pain Assessment of the Thoracic Spine and Rib Cage in Asymptomatic Individuals," *Journal of Manual & Manipulative Therapy*, vol. 16, no. 4, pp. 210-216, 2008.

- [11] K.-A. Lindgren, E. Leino and H. Manninen, "Cervical rotation lateral flexion test in brachialgia," *Archives of Physical Medicine and Rehabilitation*, vol. 73, no. 8, pp. 735-737, 1992.
- [12] J. R. Landis and G. G. Koch, "The Measurement of Observer Agreement for Categorical Data," *Biometrics*, vol. 33, no. 1, pp. 159-174, 1977.
- [13] J. Patijn, "Reproducibility and validity studies of diagnostic procedures in manual/musculoskeletal medicine," *Musculoskeletal Medicine*, no. 3, pp. 2-35, 2004.
- [14] M. T. Haneline and M. Young, "A Review of Intraexaminer and Interexaminer Reliability of Static Spinal Palpation: A Literature Synthesis," *Journal of Manipulative and Physiological Therapeutics*, vol. 32, no. 5, pp. 379-386, 2009.
- [15] J. Patijn, "Reproducibility protocol for diagnostic procedures in Manual/Musculoskeletal Medicine: Edition 2019," *Manuelle Medizin*, vol. 57, no. 6, pp. 451-479, 2019.
- [16] A. Rieder and B. Lohff, *Gender Medizin: geschlechtsspezifische Aspekte für die klinische Praxis*, Wien: Springer-Verlag, 2008.