Thieme

Intra- und Intertester-Reliabilität klinischer Tests zur Untersuchung der Bewegungskontrolle bei Patienten mit Nackenschmerzen

Systematischer Review

Intratester and Intertester Reliability of Clinical Tests for the Assessment of Movement Control in Patients with Neck Pain

Systematic Review

Autoren
Jana Allofs, Katharina van Baal, Fiona Schwarz, Katja Ehrenbrusthoff, Thomas Hering

Institute
Hochschule für Gesundheit Bochum

Schlüsselwörter
Nackenschmerzen, Reproduzierbarkeit der Ergebnisse, Assessments, Bewegungskontrolle

Key words
neck pain, reproducibility of results, assessments, movement control

eingereicht 09.03.2017
akzeptiert 14.06.2017

Bibliografie
DOI https://doi.org/10.1055/s-0044-100533
physioscience 2018; 14: 22–33

ISSN 1860-3092

Korrespondenzadresse
Jana Allofs, PT BSc
Hochschule für Gesundheit Bochum, Am Annonisbach 21, 53842 Troisdorf
jallofs@hs-gesundheit.de

ZUSAMMENFASSUNG

Hintergrund Nackenschmerzen sind weltweit eine häufig auftretende Gesundheitseinschränkung. Der große Anteil von Patienten mit unspezifischen Nackenschmerzen kann in Subgruppen eingeteilt werden. Eine dieser Subgruppen sind Patienten mit Bewegungskontrollproblemen. Für eine eindeutige Identifizierung dieser Gruppe sind zuverlässige Messverfahren notwendig. Es gibt eine Vielzahl verschiedener Tests zur Überprüfung der Bewegungskontrolle. Bisher liegt kein Review über den aktuellen Forschungsstand zur Intertester- und Intratester-Reliabilität der vielen Bewegungskontrolltests ohne technische Geräte bei Patienten mit Nackenschmerzen vor.

Ziel Das Ziel der Arbeit ist es, die aktuell vorhandene Evidenz zur Intratester- und Intertester-Reliabilität von Bewegungskontrolltests bei Patienten mit Nackenschmerzen zu untersuchen.

Methode Die Recherche fand im April 2017 bei Medline, Cochrane und PEDro unter anderem mit den Suchbegriffen „neck pain“ [Mesh], „reproducibility of results“ [Mesh], „reliability“ und „movement control impairment“ sowie einer Vielzahl von Synonymen statt. Um das Bias-Risiko der eingeschlossenen Studien zu ermitteln, wurde die QAREL-Checkliste verwendet [17]. Eine Autorin extrahierte die Studien- und Patientencharakteristika.

Ergebnisse 4 Studien mit einem geringen (8/11) bis moderaten (7/11) Bias-Risiko wurden eingeschlossen. Die Intertester-Reliabilität der 26 Tests lag zwischen ausreichend und sehr gut (k = 0,32 – 1,0), die Intratester-Reliabilität der 11 Tests war moderat bis sehr gut (k = 0,59 – 0,92). Lediglich 3 Tests wurden von je 2 Studien, alle anderen Tests jeweils nur von 1 Studie überprüft. Der am besten untersuchte und beurteilte Test zur Einschätzung der Bewegungskontrolle der HWS war der Test „Blickstabilität“.

Schlussfolgerung Weitere Studien sollten neben der Untersuchung der Validität einzelner Tests eine Testbatterie zur zuverlässigen Beurteilung der Bewegungskontrolle der HWS entwickeln.

ABSTRACT

Background Neck pain is worldwide a common health restriction. The high proportion of patients with unspecific neck pain can be allocated into subgroups. One of these is subgroups are patients with movement control problems. For a distinct identification of this group reliable assessment methods are required. There are many different tests used to examine movement control. To date there exist no reviews regarding the current state of evidence of the intertester and intratester reliability of multiple movement control tests without technical devices in patients with neck pain.

Objective The aim of this study is to evaluate the current state of evidence of the intertester and intratester reliability of movement control tests in patients with neck pain.

Method The literature research in April 2017 in Medline, Cochrane and PEDro used among others the terms "neck pain" [Mesh], "reproducibility of results" [Mesh], "reliability" and "movement control impairment" as well as multiple synonyms. In order to evaluate the included studies' risk of bias the QAREL checklist was employed [17]. One of the authors extracted the study and patient characteristics.

Results 4 studies with poor (8/11) to moderate (7/11) risk of bias were included. The intertester reliability of the 26 tests ranged from fair to very good (k = 0.32 – 1.0), the intratester reliability of the eleven tests ranged from moderate to very good (k = 0.59 – 0.92). Only 3 tests were evaluated in 2 studies, the others were merely evaluated in 1 study. The best investigated test for the evaluation of the cervical spine's movement control was the test "Blickstabilität".

Conclusion Future studies should in addition to the evaluation of the validity of single tests develop a reliable test battery for the assessment of the cervical spine's movement control.

Einleitung

Nackenschmerz ist ein häufig auftretendes Gesundheitsproblem. Die jährliche Prävalenz von Nackenschmerzen liegt je nach Studie zwischen 30 – 50 % der Bevölkerung [12, 13, 15]. Besonders Erwerbstätige leiden häufig unter Nackenschmerzen [8]. Im Jahr 2005 lag der Anteil an Erwerbstätigen mit berufsbedingten Schulter- und Nackenschmerzen in Deutschland bei 21,3 % [29].

Nackenschmerzen sind differenziert zu betrachten. Neben der zeitlichen Einteilung (akut, subakut, chronisch) werden sie auch als spezifisch oder unspezifisch klassifiziert [28]. In den meisten Fällen bleibt die Ursache der Nackenbeschwerden unklar, sodass die Mehrzahl als unspezifisch eingeordnet wird [28]. Es ist von weniger als 1 % akuter und subakuter Nackenschmerzen aufgrund einer ernsthaften Erkrankung (z. B. Tumor, Infektion) auszugehen, die damit in die Klasse der spezifischen Nackenschmerzen eingeordnet würden [28].

Gemäß einer Erhebung der European Pain Federation (EFIC) liegen allgemeine Nackenschmerzen in Europa mit 34 % auf Platz 3 der häufigsten Ursachen für chronische Schmerzen [35]. Demnach haben Nackenschmerzen unabhängig davon, ob spezifisch oder unspezifisch einen großen Anteil bei der Entstehung von chronischen Schmerzen.

Für das Gesundheitssystem stellen Beschwerden der Wirbelsäule, zu denen auch Nackenschmerzen zählen, einen enorm großen Kostenfaktor dar. Im Jahr 2008 betrugen die Kosten für Dorsopathien im Allgemeinen in Deutschland ca. 9 Milliarden Euro [33]. Genaue Angaben zu den anfallenden Kosten allein für die Behandlung der HWS liegen in Deutschland nicht vor.

Neben den gesellschaftlichen stehen die persönlichen Folgen für die Betroffenen im Vordergrund. Patienten leiden nicht nur unter den vorhandenen Schmerzen, sondern auch unter eingeschränkter Beweglichkeit der HWS, Kopfschmerzen, Schwindel, ausstrahlendem Schmerz in Schultern und Arme sowie Kraftverlust [13, 26]. Neben den direkten körperlichen Problemen kommt es zu Einschränkungen bei den Aktivitäten des täglichen Lebens. Hier klagen Patienten unter anderem über Schwierigkeiten beim Fahrrad- oder Autofahren, Computerarbeiten, Haushaltsaufgaben oder Lesen [19, 26]. Diese Beeinträchtigungen können zu einem teilweisen Rückzug der Betroffenen vom sozialen Leben und weniger Möglichkeiten der Partizipation führen. Hinzu kommen psychologische Beschwerden wie Fatigue, Stress, Frustration, Angst, Depressionen sowie die Ungewissheit und Sorgen über die Zukunft, die Arbeit und die Schwere der Erkrankung [13, 26].

Zur besseren Eingrenzung von Nackenschmerzen und deren Folgen ist ein Verständnis der Ursachen, Risikofaktoren und der Pathophysiologie von hoher Relevanz. Im Unterschied zu den bekannten Risikofaktoren für die Entstehung von Nackenschmerzen (z. B. sich wiederholende Bewegungen bei der Arbeit, eine sitzende Arbeitsposition, längere Zeiträume mit flektierter HWS, hohe psychologische Belastung im Job, psychologische Probleme, geringe Lebensqualität [6, 8, 12, 13]) ist die Pathophysiologie bei den meisten Nackenschmerzen nicht geklärt oder wird kontrovers diskutiert [13]. Um den großen Anteil an Patienten mit unspezifischen Nackenschmerzen dennoch mit geeigneten Verfahren behandeln zu können, werden sie anhand von Klassifikationssystemen Subgruppen zugeordnet, anhand derer die beste individuelle Behandlung abgeleitet werden soll [5, 24, 37, 38].

Das für die LWS entwickelte, international eingesetzte Klassifikationssystem von O'Sullivan [9, 10, 36] besitzt eine gute Reliabilität und Validität. Es ist auch auf die HWS übertragbar [24] und ordnet Patienten einer „Movement-Impairment"-(MI)- oder einer „Control-Impairment"-(CI)-Gruppe zu. CI wird als eine beeinträchtigte aktive Bewegungskontrolle während funktionellen dynamischen oder statischen Aktivitäten definiert [24]. Patienten der CI-Subgruppe zeigen lokale Schmerzen mit oder ohne Ausstrahlung, haben jedoch keine eingeschränkte Beweglichkeit in der schmerzhaften Bewegungsrichtung. Bei vorliegendem CI wird im Hinblick auf die Pathophysiologie eine Aufrechterhaltung der Schmerzen im Nacken angenommen. Dies lässt sich insbesondere auf eine veränderte neuromuskuläre Kontrolle der Nackenbewegungen zurückführen, die ungewollten Stress auf zervikale Strukturen auslöst [22].

Ein Großteil der Patienten mit unspezifischen Schmerzen soll zur CI-Gruppe zählen [24]. Daraus ergibt sich die hohe Relevanz einer zuverlässigen Identifikation der Betroffenen. Zum einen müssen Patienten mit unspezifischen Nackenschmerzen mit CI von denen ohne CI unterschieden werden können. Obwohl einige Tests zur Überprüfung der Bewegungskontrolle an der HWS vorliegen, gibt es bislang keinen Goldstandard für das sichere Erkennen von Patienten mit CI [21]. Die Eignung der Tests kann aus physiologischen Gründen angenommen werden, weil Veränderungen in der Bewegungsausführung von Patienten mit chronischen Nackenschmerzen zu beobachten sind. Diese zeigen sich in Form von Ausweichbewegungen, veränderter Statik und veränderter Bewegung der Schulter und des Schultergürtels [7, 27]. Zum anderen werden diese Verfahren sowohl in der physiothera-

peutischen Praxis und in der klinischen Forschung zur Diagnostik eingesetzt [25, 30, 31].

Neben der Unterscheidung von Personen mit von denen ohne CI ist auch die Reliabilität ein bedeutsamer Faktor bei der Beurteilung von Messinstrumenten und somit ein weiterer Aspekt der zuverlässigen Identifikation [23]. Reliabilität bezeichnet das Ausmaß der Messgenauigkeit eines Testverfahrens. Dieses ist unabhängig davon, was inhaltlich gemessen wird und meint die Wahrscheinlichkeit, mit der bei einer wiederholten Messung unter homogenen Bedingungen dasselbe Testergebnis erzielt wird. Somit ist die Reliabilität ein Maß für die Replizierbarkeit der Ergebnisse eines Tests unter gleichen Messbedingungen [26].

Therapeuten sollten bei bei 2 verschiedenen Testungen unter den gleichen Bedingungen und dem gleichen Gesundheitszustand der Patienten immer zum selben Ergebnis kommen (Intratester-Reliabilität). Außerdem müssen 2 verschiedene Therapeuten beim gleichen Patienten und unter gleichen Testbedingungen zum selben Testergebnis kommen (Intertester-Reliabilität). Nur so ist gewährleistet, dass der Test zuverlässig ist und zum richtigen Ergebnis führt [26]. Aufgrund der Relevanz von reliablen Tests im physiotherapeutischen Alltag steht dieser Aspekt im Fokus des Reviews.

Im Bereich der LWS wurden bereits einige Tests bzw. Testkombinationen zur Beurteilung der Bewegungskontrolle auf ihre Reliabilität untersucht [4, 18]. Beispielsweise fanden Luomajoki et al. [18] für 6 von 10 Tests eine Intertester-Reliabilität und für 9 von 10 eine Intratester-Reliabilität von $k > 0,6$.

Viele der im Bereich der HWS vorliegenden Tests zur Untersuchung der Bewegungskontrolle sind mit großem technischen Aufwand verbunden. Aufgrund der problematischen Praktikabilität [19] werden sie im Praxisalltag selten angewendet.

Nur wenige Studien beschäftigten sich mit praxistauglichen physiotherapeutischen Tests [25, 30]. Dies ist von besonderer Bedeutung, da Tests im klinischen Alltag und besonders in der ambulanten Praxis schnell, kostengünstig und ohne technische Geräte durchführbar sein sollten.

Im Jahr 2013 erschien ein systematischer Review unter anderem zur Reliabilität motorischer Kontrolltests der HWS, der jedoch hauptsächlich Studien zu technikbasierten Tests einschloss [20]. Im Ergebnis zeigten 2 Tests reliable Werte und werden für die Anwendung in der Praxis empfohlen: „Head Repositioning Accuracy to the Neutral Head Position“ und „The Fly“ [20].

Bisher liegt jedoch noch kein Review zur Untersuchung der Inter- und Intratester-Reliabilität von Bewegungskontrolltests der HWS ohne technische Geräte bei Patienten mit Nackenschmerzen vor. Das Ziel der vorliegenden Arbeit war es, die aktuell vorhandene Evidenz zur Reliabilität (Intratester- und Intertester-Reliabilität) von Bewegungskontrolltests ohne technische Geräte bei Patienten mit Nackenschmerzen zu untersuchen.

Methode

Für diesen anhand der PRISMA-Checkliste erstellten systematischen Literaturreview liegt kein Protokoll vor. Aufgrund der Heterogenität der Daten wurden keine Metaanalyse und keine weiteren quantitativen Analysen durchgeführt.

Ein- und Ausschlusskriterien

Patienten/Probanden

Eingeschlossen wurden Studien, die ausschließlich Patienten mit unspezifischen Nackenschmerzen untersuchten.

Alle Studien mit lediglich gesunden Probanden waren ausgeschlossen. Um Einflussfaktoren auf die Ergebnisse der Testverfahren durch andere Erkrankungen zu vermeiden, wurden außerdem Studien ausgeschlossen, die Patienten/Probanden mit radikulären Syndromen, Zustand nach Operation an der HWS oder chronischen Syndromen (z. B. Fibromyalgie) untersuchten.

Testverfahren

Eingeschlossen wurden Studien, die Bewegungskontrolltests an der HWS untersuchten. Um die kostengünstige und einfache Umsetzbarkeit in der Praxis zu gewährleisten, waren Studien ausgeschlossen, die technische Apparaturen zur Untersuchung einsetzen.

Ergebnisparameter

Eingeschlossen wurden Studien, die die Intertester- und/oder Intratester-Reliabilität von Testverfahren zur Messung der Bewegungskontrolle der HWS bei Patienten mit Nackenschmerzen untersuchten. Bei der Entwicklung neuer Testverfahren wird vor der Validität in der Regel zuerst die Reliabilität untersucht, da ein Test nicht als valide gilt, wenn die Testergebnisse sich als unzuverlässig herausstellen [1]. Daher lag der Fokus des vorliegenden Reviews zunächst auf der Betrachtung der Intertester- und Intratester-Reliabilität.

Studientypen

Eingeschlossen wurden Reliabilitäts-, Reproduzierbarkeits- und Querschnittsstudien mit Fokus auf der Überprüfung der Reliabilität von Messverfahren. Für eine hohe Aussagekraft des Reviews waren Fallberichte und Einzelfallstudien ausgeschlossen.

Qualität der Studien

Eingeschlossen wurden Studien, die eine methodische Qualität von ≥ 7 Punkten in der QAREL-Checkliste erreichten. Studien mit einem Score von < 7 entsprächen einem hohen Bias-Risiko [32] und waren zugunsten einer verbesserten Aussagekraft des Reviews ausgeschlossen.

Literaturrecherche

Datenbanken

Im April 2017 wurde eine systematische Literaturrecherche in den Datenbanken Medline (PubMed), PEDro sowie The Cochrane Library durchgeführt und somit davor veröffentlichte Studien eingeschlossen.

Suchbegriffe und Suchstrategie

Unter anderem wurden die Suchbegriffe „neck pain“ [Mesh], „reproducibility of results“ [Mesh], „reliability“ und „movement control impairment“, synonyme Schlagworte und Stichworte sowie

vorhandene Mesh-Begriffe verwendet. Zur Verringerung des Search-Bias erstellten 3 Autorinnen die Liste der Suchbegriffe und änderten sie je nach Konsens ab. Die abschließende Suche fand am 15.04.2017 statt. Eingeschlossen wurden auf Englisch und Deutsch publizierte Studien, ohne weitere Limitierungen. Die Handsuche ergab keine zusätzlichen Treffer (▶ **Tab. 1**).

Studienauswahl

Im nächsten Schritt wurden alle nicht den Einschlusskriterien entsprechende Arbeiten und Duplikate aus den verschiedenen Datenbanken ausgeschlossen. Die Studienauswahl erfolgte in 2 Schritten. Zunächst wurde durch Screening des Titels und des Abstracts aller Treffer eine erste Auswahl getroffen und anschließend anhand des Volltextes über die Aufnahme in den Review entschieden. 3 Autorinnen führten jeweils unabhängig voneinander alle Teilschritte durch und trafen im Konsens eine endgültige Auswahl relevanter Studien.

Bewertung des Bias-Risikos und der methodischen Qualität der Studien

Zur kritischen Prüfung der methodischen Qualität der eingeschlossenen Studien diente die QAREL-Checkliste. Bereits mehrere Reviews untersuchten die Reliabilität diagnostischer Tests [3, 4, 32]. QAREL wurde in Anlehnung an die „Standards for Reporting Studies of Diagnostic Accuracy“ (STARD; [2] und das „Quality Assessment of Diagnostic Accuracy Studies“ (QUADAS; [39, 40]) entwickelt. Da bislang keine validierte deutsche Version der QAREL-Checkliste vorliegt, kam für den vorliegenden Review die englische Version zum Einsatz. Die Checkliste umfasst 11 Elemente, wie z. B. Grundsätze zu Verblindung, Pausendauer zwischen den einzelnen Messungen, Testdurchführung und -interpretation sowie zur statistischen Auswertung. Die Items wurden jeweils separat bezüglich ihrer Bedeutsamkeit für die Qualität der Studie eingeschätzt [17]. Jedes mit „ja“ beantwortete Item wurde mit „1“ codiert, sodass sich eine mögliche Gesamtpunktzahl von maximal 11 ergab. Eine höhere Punktzahl steht für eine bessere methodische Qualität. Werte < 7 bedeuteten ein hohes, 7 ein moderates und ≥ 8 ein niedriges Bias-Risiko [32]. 3 Autorinnen bewerteten unabhängig voneinander die 4 eingeschlossenen Studien mithilfe von QAREL. Anschließend verglichen und diskutierten sie die Ergebnisse, um schließlich einen Konsens zu finden.

Datenanalyse

Mit Blick auf die Heterogenität der in den eingeschlossenen Studien untersuchten Tests und der Studienpopulationen wurden die Ergebnisse mithilfe deskriptiver Statistiken zusammengefasst. Eine Autorin extrahierte die Studien- und Patientencharakteristiken (Studientyp, Studienteilnehmer, untersuchte Testverfahren, Charakteristika der Tester, Endpunkte) sowie die Hauptergebnisse der einzelnen Studien.

Zur Beurteilung der Reliabilität der klinischen Messinstrumente wurde der Cohens Kappa-Koeffizient zwischen den Studien verglichen. Im vorliegenden Review wurde die Übereinstimmung bei Werten < 0,20 als ungenügend, 0,21 – 0,40 als ausreichend, 0,41 – 0,60 als moderat, 0,61 – 0,80 als gut und 0,81 – 1,0 als sehr gut bewertet [16].

▶ **Tab. 1** Suchstrategie in den Datenbanken Medline und Cochrane.

	Suchbegriffe	Treffer	
		Medline	Cochrane
#1	chronic neck pain	3789	908
#2	chronic non-specific neck pain	116	68
#3	neck pain	24 055	3431
#4	chronic cervical pain	4836	395
#5	chronic non-specific cervical pain	125	19
#6	cervical pain	33 771	2445
#7	“neck pain” [Mesh]	5519	786
#8	movement control impairment	4356	830
#9	movement control dysfunction	30 188	632
#10	movement system impairment	2801	211
#11	motor control impairment	5317	1508
#12	relative flexibility	6457	131
#13	motor control test	18 435	5544
#14	movement control test	30 518	3876
#15	reliability	129 120	6900
#16	test-retest reliability	353 235	1059
#17	test retest reliability	353 235	1112
#18	intertester reliability	264	21
#19	inter tester reliability	168	19
#20	intratester reliability	209	11
#21	intra tester reliability	148	22
#22	“reproducibility of results” [Mesh]	330 992	11 267
#23	#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7	33 771	4914
#24	#8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14	73 140	9725
#25	#15 OR #16 OR #17 OR #18 #OR #19 OR #20 OR #21 OR #22	425 204	16 394
#26	#23 AND #24 AND #25	32	8

Für eine Gesamtaussage zur Evidenz der einzelnen Tests wurden die Ergebnisse der einzelnen Studien in Form einer qualitativen Analyse nach den COSMIN-Guidelines zusammengeführt [34]. Die Einteilung erfolgte in 5 verschiedene Evidenzlevel (stark, moderat, limitiert, widersprüchlich und unbekannt) in Abhängigkeit von der methodischen Qualität der Studien und nach der Anzahl an Studien, die den jeweiligen Test untersuchten (▶ **Tab. 2**).

▶ **Tab. 2** Evidenzlevel für die Qualität der Eigenschaften der Messverfahren gemäß der COSMIN-Guideline [34].

Level	Bewertung	Kriterien
stark	+++ oder – – –	einheitliche Ergebnisse in mehreren Studien mit guter methodischer Qualität ODER in 1 Studie mit exzellenter methodischer Qualität
moderat	++ oder – –	einheitliche Ergebnisse in mehreren Studien mit ausreichender methodischer Qualität ODER in 1 Studie mit guter methodischer Qualität
limitiert	+ oder –	1 Studie mit ausreichender methodischer Qualität
widersprüchlich	±	widersprüchliche Ergebnisse
unbekannt	?	nur Studien mit unzureichender methodischer Qualität

+ = positiv; ? = unbekannt; – = negativ.

Die Aussagekraft der Studien mit erhöhtem Bias-Risiko bzw. geringerer methodischer Qualität wurde niedriger eingestuft und somit bei der Bestimmung des Evidenzlevels berücksichtigt [34].

Ergebnisse

Ergebnisse der Recherche

Nach Entfernung der Duplikate ergaben sich 35 verschiedene Treffer, von denen nach einer ersten Bewertung aus verschiedenen Gründen 27 ausgeschlossen wurden (▶ **Abb. 1**). Von den verbliebenen 8 Studien wurden die Volltexte auf Eignung geprüft. Insgesamt erfüllten 4 der 8 Studien die Einschlusskriterien und fanden Berücksichtigung im Review [11, 14, 25, 30].

Methodische Qualität der eingeschlossenen Studien

Der Konsensus der unabhängigen Bewertung der 3 Autorinnen der mithilfe der QAREL-Checkliste analysierten Studien ist in ▶ **Tab. 3** dargestellt. 3 Studien [11, 14, 30] wiesen ein geringes (8/11) und 1 Studie [25] ein moderates Bias-Risiko (7/11) auf. Besonders die Verblindung der Untersucher gegenüber klinischen Informationen [14, 25] und speziellen Kennzeichen der Patienten wie Narben, Tattoos oder Ähnliches [11, 14, 25, 30] waren zum Teil nicht erfüllt oder unklar.

Studien- und Patientencharakteristika

Die 4 in den Review eingeschlossenen Studien umfassten insgesamt n = 140 Patienten mit Nackenschmerzen. Je nach Studie lag die Teilnehmeranzahl zwischen 21 und 45.

Einschlusskriterien aller Studien waren Schmerzen im Bereich der HWS und des Nackens, die bei 2 Studien länger als 3 Monate [11, 30] und bei 1 Studie länger als 6 Monate [14] bestehen mussten.

Die Ausschlusskriterien der 4 eingeschlossenen Studien ähnelten sich: schwere neurologische Einschränkungen oder zentralneurologische Erkrankungen, schwerwiegende Grunderkrankungen wie Tumore, Augenerkrankungen und andere chronische Erkrankungen wie Fibromyalgie, ein vor kurzer Zeit erlittenes Trauma [11], eine vorliegende Fraktur [14] oder eine Operation der HWS [25, 30].

Alle Studien analysierten die Intertester-Reliabilität, Segarra et al. [30] und Jorgensen et al. [14] zusätzlich die Intratester-Reliabilität (▶ **Tab. 4, 5**). Alle Studien zusammen untersuchten 11 verschiedene nicht technikbasierte Tests auf ihre Intratester-Reliabilität. Insgesamt 26 nicht technikbasierte Tests wurden auf ihre Intertester-Reliabilität, davon 3 jeweils von 2 Studien überprüft.

Die in den 4 Studien untersuchten Tests waren sehr unterschiedlich. Della Casa et al. [11] und Jorgensen et al. [14] beurteilten z. B. nicht nur die Bewegungskontrolle der HWS, der Schultern und des Kopfes, sondern auch Augenbewegungen und die Blickstabilität. Die anderen Studien bewerteten die Bewegungsqualität bzw. Ausweichbewegungen bei spezifischen Kopf- und HWS-Bewegungen [25, 30].

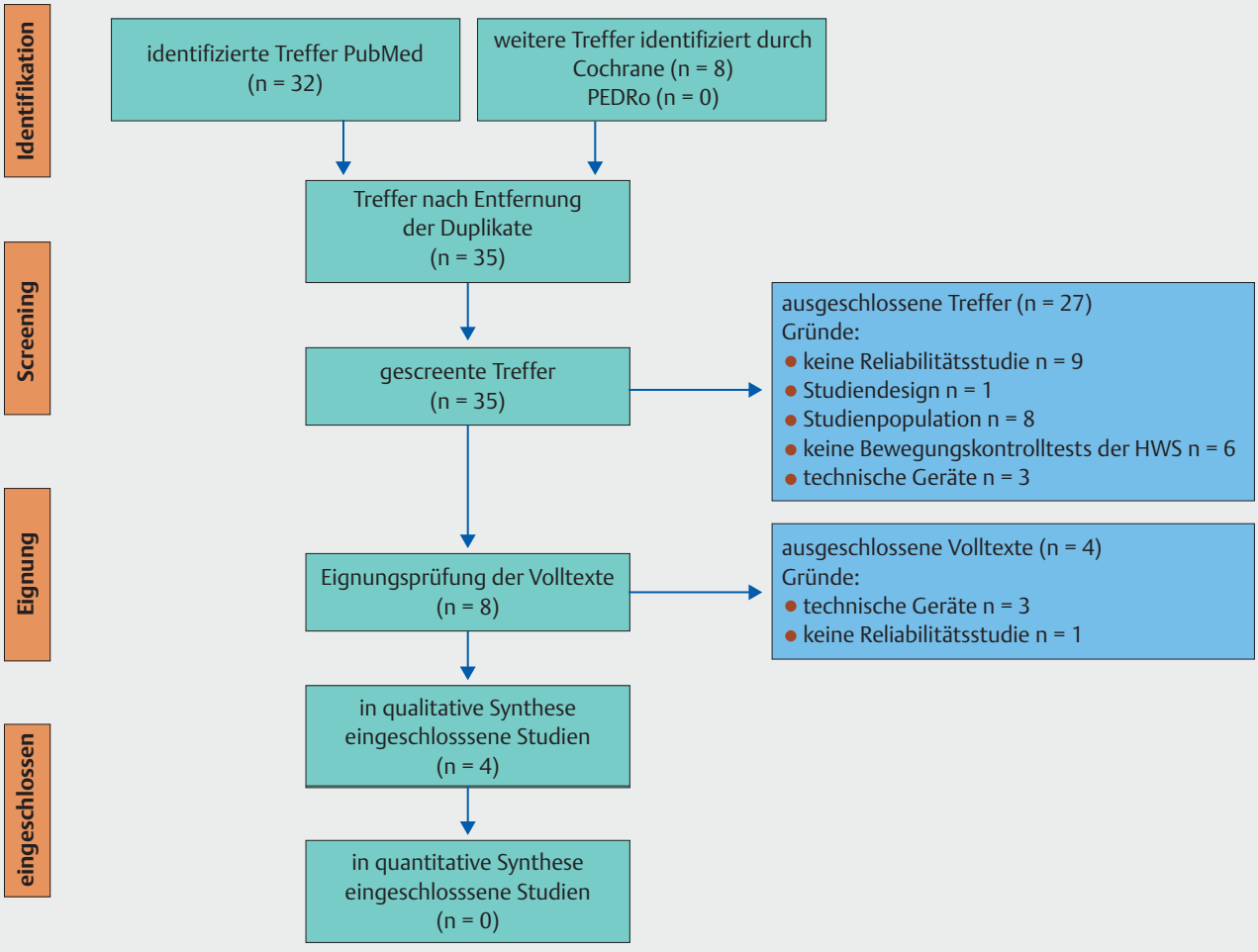
Die meisten Tester waren berufserfahren, was jedoch nur Segarra et al. [30] genauer definierten. Die Autoren verglichen einen Therapeuten mit 1 Jahr Berufserfahrung mit einem Physiotherapeuten mit 10 Jahren Berufserfahrung.

Zusammenfassung der Ergebnisse der Studien

Insgesamt untersuchten die Studien 26 verschiedene Tests, von denen 3 von jeweils 2 Studien beurteilt wurden (aktive unilaterale Armflexion, aktive HWS-Rotation im Sitz, Blickstabilität; [25, 30]). Jorgensen et al. [14] beschäftigten sich mit 2 Tests zu 2 verschiedenen Messzeitpunkten (Blickstabilität, Smooth Pursuit Neck Torsion Test). Die anderen Tests wurden nur in einer einzigen Studie bewertet. Alle Tests wiesen insgesamt eine sehr gute bis ausreichende Intertester-Reliabilität auf.

Die 31-mal beurteilte Intertester-Reliabilität erzielte 9-mal eine sehr gute (Kappa [k] = 0,81 – 1,0), 16-mal eine gute (k = 0,66 – 0,80), 4-mal eine moderate (k = 0,41 – 0,60) und 2-mal eine ausreichende Übereinstimmung (k = 0,21 – 0,40). Es ist jedoch zu beachten, dass die 95 %-Konfidenzintervalle der einzelnen Kappa-Werte eine große Spannweite aufwiesen. So liegen die 95 %-Konfidenzintervalle der als sehr gut beurteilten Tests zwischen moderat und sehr gut, die der als gut beurteilten Tests zwischen ausreichend und sehr gut. Die Konfidenzintervalle der als moderat oder ausreichend beurteilten Tests nehmen die gesamte Spannweite von ungenügend bis sehr gut ein.

Die Untersuchung der Intratester-Reliabilität erfolgte bei 11 Tests. Jeder Test wurde nur von einer einzigen Studie überprüft. Innerhalb dieser Studien nahmen jedoch 2 Tester die Beurteilung vor. Davon erreichten 8 Untersucher eine sehr gute (k = 0,81 – 0,92), 13 Untersucher gute (k = 0,70 – 0,80) und 1 Untersucher eine moderate (k = 0,59) Intratester-Reliabilität. Auch hier ist die relativ große Spannweite der 95 %-Konfidenzintervalle (moderat



▶ **Abb. 1** Flowchart zu Literaturrecherche und Studienauswahl. (Quelle: J. Allofs; graf. Umsetzung: Thieme Gruppe).

bis sehr gut) der als gut oder sehr gut beurteilten Tests zu beachten. Der als moderat beurteilte Test weist ein 95 %-Konfidenzintervall mit Werten von ausreichend bis sehr gut auf.

Die beste Intratester-Reliabilität für beide Untersucher erzielte der Test „aktive unilaterale Armflexion im Stand“ ([30]; A: k = 0,90; 0,63 – 1,0); B: k = 0,89; 0,69 – 1,0). Ebenfalls sehr gute Werte erreichten „aktive HWS-Extension im Vierfüßlerstand“ ([30]; A: k = 0,86; 0,66 – 1,0; B: k = 0,84; 0,53 – 1,0) und „Zurückführen des Kopfes aus maximaler Extension in die Neutralposition im Sitz“ ([30]; A: k = 0,87; 0,68 – 1,0; B: k = 0,85; 0,64 – 1,0).

Die beste Intertester-Reliabilität wiesen „bilaterale Schulterelevation“ und „Vorbeugen im Stand“ mit jeweils k = 1,0 sowie „Pro-Retraktion“ mit k = 0,91 (0,75 – 1,0) auf [25].

Beurteilung der methodischen Qualität der Tests

Da die Studien eine weitgehend einheitliche methodische Qualität aufwiesen, musste lediglich eine geringfügige Gewichtung der Ergebnisse erfolgen. Bei der qualitativen Analyse fand die QAREL-Beurteilung der Studie von Patroncini et al. ([25]; moderates Bias-Risiko) Berücksichtigung, weil sie im Vergleich zu den anderen Studien eine etwas schlechtere methodische Qualität erzielte. Daher war von einem größeren Bias-Risiko auszugehen.

Insgesamt beurteilten je 2 Studien die Intertester-Reliabilität von 3 Tests [11, 14, 25, 30]. Patroncini et al. [25] und Segarra et al. [30] untersuchten „aktive unilaterale Armflexion“ sowie „aktive HWS-Rotation im Sitz“, Della Casa et al. [11] und Jorgensen et al. [14] „Blickstabilität“.

Bei „Blickstabilität“ errechneten Della Casa et al. [11] einen Weighted-Cohen's-Kappa-Wert (wk) von wk = 0,86 und Jorgensen et al. [14] k = 0,71 bzw. k = 0,66. Damit zeigten 2 Studien von guter Qualität eine starke Evidenz, dass „Blickqualität“ eine gute bis sehr gute Intertester-Reliabilität aufweist.

Dahingegen fanden sich uneinheitliche Aussagen hinsichtlich der Ergebnisse zu „aktive unilaterale Armflexion“ und „aktive HWS-Rotation im Sitz“. Für „aktive unilaterale Armflexion“ ermittelten Segarra et al. [30] einen Wert von k = 0,32 im Gegensatz zu Patroncini et al. [25] mit k = 0,74. Zu „aktive unilaterale Armflexion“ ergab sich widersprüchliche Evidenz aus 2 Studien mit moderater bis guter Qualität, inwieweit der Test eine ausreichende oder gute Intertester-Reliabilität aufweist.

▶ **Tab. 3** Bewertung der methodischen Qualität der eingeschlossenen Studien anhand der QAREL-Checkliste [17].

QAREL-Items	Segarra et al. [30]	Della Casa et al. [11]	Patrocini et al. [25]	Jorgensen et al. [14]
1	j (571)[1]	j (3)[1]	j (3)[1]	j (2)[1]
2	j (571)[1]	j (6)[1]	j (3)[1]	j (2)[1]
3	j (573)[1]	j (6)[1]	j (3)[1]	j (2)[1]
4	j (573)[1]	n/a[2]	n/a[1]	j (2)[2]
5	n/a[1]	n/a[1]	n/a[1]	n/a[1]
6	j (573)[1]	j (6)[1]	n[2]	n[2]
7	u[1]	u[1]	u[1]	u[1]
8	u[2]	j(6)[1]	j (3)[1]	j (5)[1]
9	j (573)[2]	j (6)[1]	j (3)[1]	j (5)[1]
10	j (572/573)[1]	j (3 – 6)[1]	j (2 – 3)[1]	j (3 – 5)[1]
11	j (573)[1]	j (6)[1]	j (4 – 5)[1]	j (5)[1]
Summe	8	8	7	8
Bias-Risiko	gering	gering	moderat	gering

() = Seitenzahl in der Studie, auf der der Parameter berichtet wird; j = ja; n = nein; n/a = Item nicht zutreffend für die Studie; u = unklar. 1. Wurden die Tests in einer Probandengruppe untersucht, die repräsentativ für das von den Autoren beschriebene Krankheitsbild war? 2. Entsprachen die Untersucher der Art von Untersuchern, für die die Tests entwickelt wurden? 3. Waren die Untersucher in Hinsicht auf die Ergebnisse untereinander verblindet? 4. Waren die Untersucher in Hinsicht auf ihre ersten Ergebnisse verblindet? 5. Waren die Untersucher in Hinsicht auf das Ergebnis des Referenzstandards (Goldstandard) verblindet? 6. Waren die Untersucher in Hinsicht auf für die Studie nicht relevante klinische Informationen verblindet? 7. Waren die Untersucher gegenüber speziellen Kennzeichen wie Tattoos, Narben und Akzent verblindet? 8. War die Reihenfolge der Untersuchung randomisiert? 9. War das Zeitintervall zwischen den durchgeführten Tests angemessen, um eine stabile Symptomatik/stabile zu messende Variable zu gewährleisten? 10. Wurde der Test richtig und angemessen durchgeführt? 11. Wurden die passenden statistischen Verfahren bei der Auswertung des Testes verwendet?
[1] alle Rater stimmten direkt überein.
[2] Rater stimmten nicht direkt überein, Konsensus wurde gefunden.

Bei „aktive HWS-Rotation im Sitz“ stellten Segarra et al. [30] einen k-Wert von 0,81, Patroncini et al. [25] jedoch nur 0,47 fest. Zu „aktive HWS-Rotation im Sitz“ zeigten somit 2 Studien mit moderater bis guter Qualität widersprüchliche Evidenz, inwieweit der Test moderate oder sehr gute Intertester-Reliabilität besitzt.

Aufgrund der moderaten und damit schlechteren methodischen Qualität bei Patroncini et al. [25] war tendenziell eine höhere Aussagekraft und ein geringeres Bias-Risiko der Ergebnisse von Segarra et al. [30] zu erwarten. Dies führte jedoch zu keiner Änderung des Evidenzlevels. Beide Tests sollten in Zukunft weiter überprüft werden, um die widersprüchlichen Ergebnisse besser beurteilen zu können. Bisher können beide Tests noch nicht eindeutig empfohlen werden.

Ausgehend von Segarra et al. [30] bestand kein erheblicher Einfluss der Berufserfahrung der Untersucher auf die Intratester-Reliabilität der Bewegungskontrolltests für die HWS, Die Werte der beiden Untersucher lagen stets sehr nah beieinander (▶ **Tab. 5**).

Diskussion

Dieser systematische Review untersuchte als erster die Reliabilität von Bewegungskontrolltests ohne den Einsatz technischer Geräte an der HWS bei Patienten mit Nackenschmerzen.

Hauptergebnisse

Die eingeschlossenen Studien wiesen ein moderates bis geringes Bias-Risiko auf und spiegelten die große Anzahl an verfügbaren Bewegungskontrolltests an der HWS wider. Daher wurden nur wenige Tests in mehreren Studien untersucht. Alle eingeschlossenen Studien überprüften die Intertester- [11, 14, 25, 30], davon 2 zusätzlich die Intratester-Reliabilität [14, 30].

Die Intertester-Reliabilität lässt sich zusammenfassend besser bewerten, da sie von mehreren Studien untersucht wurde. Eine Gesamtzahl von 3 Tests wurde von jeweils 2 Studien beurteilt [11, 14, 25, 30]. „Blickstabilität“ erzielte als einziger Test gute bis sehr gute Werte für die Intertester-Reliabilität ([11, 14]; k = 0,66 – 0,86). Dahingehend stimmten die Ergebnisse der beiden Studien jeweils überein [11, 14].

Die beiden anderen zweifach untersuchten Tests (aktive HWS-Rotation im Sitz und aktive unilaterale Armflexion [25, 30]) wurden unterschiedlich beurteilt. Segarra et al. [30] bewerteten „aktive HWS-Rotation im Sitz“ als sehr gut (k = 0,81), Patroncini et al. [25] dagegen als moderat (k = 0,47). „Aktive unilaterale Armflexion“ bewerteten Segarra et al. [30] als ausreichend (k = 0,32) und Patroncini et al. [25] als gut (k = 0,74). Zudem wiesen die 95 %-Konfidenzintervalle der genannten Kappa-Werte der 2 Tests in beiden Studien eine große Spannweite auf, was die sichere Beurteilung ihrer Reliabilität erschwerte (▶ **Tab. 2, 5**). Eine Ursache für die uneinheitlichen Ergebnisse könnte die unter-

▶ **Tab. 4** Intratester-Reliabilität der Bewegungskontrolltests für die HWS.

Autor	Test	Intratester-Reliabilität k (95 %-KI)	Gesamtbeurteilung der Reliabilität
Segarra et al.[30]	aktive HWS-Extension im Vierfüßlerstand	A: k = 0,86 (0,66 – 1,0) B*: k = 0,84 (0,53 – 1,0)	sehr gut sehr gut
	HWS-Rotation im Vierfüßlerstand	A: k = 0,80 (0,55 – 1,0) B*: k = 0,79 (0,54 – 0,99)	gut gut
	aktive HWS-Flexion im Vierfüßlerstand	A: k = 0,70 (0,40 – 0,92) B*: k = 0,70 (0,45 – 0,93)	gut gut
	aktive HWS-Extension im Sitzen	A: k = 0,74 (0,49 – 0,94) B*: k = 0,71 (0,44 – 0,93)	gut gut
	Rückführen in Nullstellung aus Extension im Sitzen	A: k = 0,87 (0,68 – 1,0) B*: k = 0,85 (0,64 – 1,0)	sehr gut sehr gut
	aktive bilaterale Armflexion im Stand	A: k = 0,78 (0,53 – 0,98) B*: k = 0,77 (0,52 – 0,97)	gut gut
	Nach-hinten-Setzen im Vierfüßlerstand	A: k = 0,80 (0,54 – 1,0) B*: k = 0,78 (0,54 – 0,99)	gut gut
	aktive unilaterale Armflexion im Stand	A: k = 0,90 (0,63 – 1,0) B*: k = 0,89 (0,69 – 1,0)	sehr gut sehr gut
	aktive HWS-Rotation im Sitz	A: k = 0,81 (0,55 – 1,0) B*: k = 0,80 (0,55 – 1,0)	sehr gut gut
Jorgensen et al. [14]	Blickstabilität	A*: k = 0,92 (0,78 – 1,0) B*: k = 0,78 (0,54 – 1,0)	sehr gut gut
	Smooth Persuit Neck Torsion Test	A*: k = 0,59 (0,24 – 0,93) B*: k = 0,74 (0,47 – 1,0)	moderat gut

95 %-KI = 95 %-Konfidenzintervall; A = Untersucher 1; B = Untersucher 2; A*/B* = Untersucher Berufsanfänger/Studenten; k = Kappa-Wert.

schiedliche Methodik der Studien sein, auf die später noch genauer eingegangen wird. Die Ergebnisse von Segarra et al. [30] waren aufgrund der vorliegenden Verblindung der Untersucher und dem insgesamt geringeren Bias-Risiko aussagekräftiger.

Insgesamt wurden 26 Tests auf ihre Intertester-Reliabilität hin untersucht, davon 3 Tests jeweils von 2 Studien [11, 14, 25, 30] sowie 2 Tests in 1 Studie zu 2 Messzeitpunkten [14]. Dabei schnitten 9 Tests sehr gut, 16 gut, 4 moderat und 2 ausreichend ab.

Die Intratester-Reliabilität von 11 Tests wurde in 2 Studien von 2 Untersuchern beurteilt [14, 30]. Hier wiesen 8 Tests eine sehr gute, 13 eine gute und 1 Test eine moderate Intratester-Reliabilität auf.

Insgesamt ist jedoch kritisch zu betrachten, dass die 95 %-Konfidenzintervalle der einzelnen Kappa-Werte über eine große Spannweite verfügten (▶ **Tab. 5**). Daher lassen sich die Angaben zur Zuverlässigkeit nur mit einiger Unsicherheit verallgemeinern. Die Zusammenfassung der beurteilten Tests gibt somit nur einen groben Überblick.

Stärken der Studien

Bei der methodischen Beurteilung mithilfe der QAREL-Checkliste ergab sich ein moderates (7/11; [25]) bis geringes (8/11; [11, 14, 30]) Bias-Risiko. Diese insgesamt gute methodische Qualität wird zusätzlich durch den Umstand aufgewertet, dass einige Kriterien der QAREL-Checkliste für manche der beurteilten Studien nicht relevant waren (z. B. Verblindung gegenüber dem Referenztest, da kein Referenztest vorliegt).

Limitationen der Studienergebnisse

Die Aussagekraft der Ergebnisse ist insbesondere deshalb kritisch zu beurteilen, weil der Großteil der Tests lediglich in einer einzigen Studie untersucht wurde [30]. Dies stellt auch eine wichtige Limitation in der Methodik der eingeschlossenen Studien dar. Die Testauswahl in den einzelnen Studien ist teilweise nicht hinreichend begründet, was zur großen Anzahl an beurteilten Tests führte und eine umfassende Aussage zu deren Reliabilität erschwerte.

Im Hinblick auf die visuelle Beobachtung der Patienten bei den Tests ist weiterhin die lediglich qualitative Testung der Bewegungskontrolle in allen eingeschlossenen Studien zu kritisieren, was einen hohen Grad an Subjektivität bedeutet.

Problematisch ist ferner das Training der Tester in den eingeschlossenen Studien. Es ist naheliegend, dass ein unterschiedlich langes und ausgeprägtes Training eine Rolle bei der Beurteilung der Testverfahren und damit für die Reliabilität gespielt haben kann. Bei Segarra et al. [30] erhielten die Untersucher vor Studienbeginn 3-mal 1 Stunde Training. Bei Jorgensen et al. [14] erklärte ein erfahrener Physiotherapeut den Untersuchern die Tests zunächst detailliert, im Anschluss wurden 10 Beispielpatienten in einer offenen Runde getestet und die Ergebnisse gemeinsam

▶ **Tab. 5** Intertester-Reliabilität der Bewegungskontrolltests für die HWS.

Autor	Test	Intertester-Reliabilität k, wk (95 %-KI)	Gesamtbeurteilung der Reliabilität
Segarra et al. [30]	aktive HWS-Extension im Vierfüßlerstand	k = 0,67 (0,33 – 0,95)	gut
	HWS-Rotation im Vierfüßlerstand	k = 0,80 (0,66 – 0,93)	gut
	aktive HWS-Flexion im Vierfüßlerstand	k = 0,52 (0,22 – 0,81)	moderat
	aktive HWS-Extension im Sitzen	k = 0,73 (0,29 – 0,91)	gut
	Rückführen in Nullstellung aus Extension im Sitzen	k = 0,69 (0,44 – 0,90)	gut
	aktive bilaterale Armflexion im Stand	k = 0,71 (0,44 – 0,93)	gut
	Nach-hinten-Setzen im Vierfüßlerstand	k = 0,36 (0,21 – 0,68)	ausreichend
Segarra et al. [30] Patroncini et al. [25]	aktive unilaterale Armflexion	k = 0,32 (0,08 – 0,63) k = 0,74 (0,47 – 0,95)	ausreichend gut
	aktive HWS-Rotation im Sitz	k = 0,81 (0,58 – 1,0) k = 0,47 (0,04 – 0,89)	sehr gut moderat
Della Casa et al. [11] Jorgensen et al. [14]	Blickstabilität	wk = 0,86 (0,75 – 0,97) T1: k = 0,71 (0,50 – 0,93) T2: k = 0,66 (0,42 – 0,91)	sehr gut gut gut
Della Casa et al. [11]	sequenzielle Kopf- u. Augenbewegungen	wk = 0,86 (0,76 – 0,97)	sehr gut
	Augenbewegung in 45° relativer Nackendrehung nach rechts	wk = 0,54 (0,29 – 0,79)	moderat
	Augenbewegung in 45° relativer Nackendrehung nach links	wk = 0,79 (0,62 – 0,97)	gut
	Augenbewegung neutral	wk = 0,72 (0,55 – 0,88)	gut
Patroncini et al. [25]	bilaterale Schulterelevation	k = 1,0	sehr gut
	Lateralflexion	k = 0,77 (0,55 – 0,97)	gut
	Extension mit Retraktion	k = 0,68 (0,47 – 0,9)	gut
	Nicken an der Wand	k = 0,80 (0,55 – 1,0)	gut
	Lateralflexion + kontralaterale Rotation	k = 0,68 (0,47 – 0,89)	gut
	komplette Flexion u. Extension	k = 0,69 (0,47 – 0,90)	gut
	Oberkörper nach vorne u. hinten	k = 0,84 (0,68 – 0,94)	sehr gut
	Armflexion bis 90° mit Gewicht	k = 0,85 (0,55 – 1,0)	sehr gut
	Vorbeugen im Stand	k = 1,0	sehr gut
	Flexion in Rückenlage	k = 0,81 (0,61 – 1,0)	sehr gut
	Proretraktion	k = 0,91 (0,75 – 1,0)	sehr gut
Jorgensen et al. [14]	Smooth Pursuit Neck Torsion Test	t1: k = 0,46 (0,17 – 0,76) t2: k = 0,71 (0,47 – 0,94)	moderat gut

95 %-KI = 95 %-Konfidenzintervall; k = Kappa-Wert; t1 = Messzeitpunkt 1; t2 = Messzeitpunkt 2; wk = Weighted Cohen's Kappa.

interpretiert. Im Gegensatz dazu war das Training mit der Betrachtung zweier Patientenbeispiele bei Patroncini et al. [25] deutlich weniger umfangreich. Bei Della Casa et al. [11] erhielten die Untersucher außer einem kurzen Instruktionsvideo kein spezifisches Training. Möglicherweise müssten für die unterschiedlichen Testverfahren differenziert Art und Umfang des Trainings der Untersucher abgestimmt werden.

Eine weitere Schwierigkeit beim Vergleich der Studienergebnisse stellen die verschiedenen Patientenpopulationen dar. Neben den Einschluss- unterschieden sich auch die Ausschlusskriterien teilweise erheblich.

Den methodisch wichtigen Aspekt der Verblindung brachten nicht alle Studien gleich ein. Bei Patroncini et al. [25] und Jorgensen et al. [14] offenbarten sich Mängel bei der Verblindung: Im einen Fall war nur einer der beiden Untersucher bezüglich der Baseline-Daten der Probanden verblindet [25], im anderen waren von 3 Untersuchern lediglich 2 ausreichend verblindet [14]. Dies kann ein Grund für die besseren Kappa-Werte in den beiden genannten Studien gewesen sein, da eine mangelhafte Verblindung

die Ergebnisse verbessert haben könnte. Eine unzureichende Verblindung birgt die Gefahr, dass den Testern möglicherweise die Patienten und deren klinische Symptomatik bereits bekannt sind und die untersuchten Tests somit nicht neutral überprüft wurden. Bei der dahingehenden Betrachtung der Testergebnisse fällt auf, dass von den in mehreren Studien untersuchten Tests „aktive HWS-Rotation im Sitz“ und „Blickstabilität“ keine besseren Werte aufwiesen, „aktive unilaterale Armflexion“ allerdings schon [25].

Die 4 Studien untersuchten zwar in ausreichendem Maße zumindest eine Form der Reliabilität (Intratester- und/oder Intertester-Reliabilität) ihrer Tests, beurteilten aber kaum die Validität. So untersuchten die meisten Studien z. B. nicht oder nicht ausreichend die Kriteriums- und Konstruktvalidität. Dies ist auf den aktuell nicht vorhandenen Goldstandard zur Beurteilung der Bewegungskontrolle bei Patienten mit Nackenschmerzen zurückzuführen [21].

Vergleich der Ergebnisse mit anderen Studien

Im Vergleich der Ergebnisse des vorliegenden Reviews mit anderen Studien bzw. systematischen Übersichtsarbeiten finden sich für technikbasierte Bewegungskontrolltests ebenfalls gute Reliabilitäts- und Validitätskennwerte [20]. Die methodische Qualität der im Review von Michiels et al. [20] eingeschlossenen Studien war genau wie im vorliegenden Review im Allgemeinen gut. Obwohl der Test „The Fly“ mit einer moderaten bis sehr guten Test-Retest-Reliabilität (Intraklassen-Korrelationskoeffizient [ICC] = 0,60 – 0,86) ein besseres Ergebnis aufwies als der „Head Repositioning Accuracy to the Neutral Head Position“ mit einer ausreichenden bis sehr guten Test-Retest-Reliabilität (ICC = 0,35 – 0,87), werden beide zur Testung der motorischen Kontrolle empfohlen. Vergleichbar mit den Ergebnissen des vorliegenden Reviews ist trotz der zahlreichen untersuchten Tests nur „Blickstabilität“ (k = 0,66 – 0,71; wk = 0,86) zu empfehlen. Da die im vorliegenden Review ebenfalls als gut oder sehr gut bewerteten Tests nur von 1 Studie untersucht wurden, scheint eine weitere Überprüfung sinnvoll.

Ein systematischer Review zu Bewegungskontrolltests der LWS aus 2013 [4] kommt hinsichtlich der Anzahl vorhandener Bewegungskontrolltests zu einem ähnlichen Ergebnis wie der vorliegende Artikel. Für die LWS wurden 19 verschiedene Tests und eine Testbatterie aus 8 Studien identifiziert. Diese große Anzahl entspricht der von 26 Tests für die HWS. Wie bei der vorliegenden Arbeit und dem systematischen Review zu motorischen Kontrolltests der HWS [20] sind auch für die LWS nur wenige Tests empfehlenswert (One Leg Stance, k = 0,43 – 1,9 und Prone Knee Flexion, k = 0,43 – 0,76). Mehrere Studien mit geringem Bias-Risiko haben diese beiden Tests getestet und mit moderat bis sehr gut bewertet [4].

Stärken des Reviews

Insgesamt ist durch die unabhängige Konsensfindung im Autorenteam bei der Auswahl relevanter Studien die Wahrscheinlichkeit eines Search-Bias als gering einzuschätzen. Auch die unabhängige Bewertung der Studienmethodik und die interne Absprache unter den Untersucherinnen zur Anwendung der QAREL-Checkliste mit Absprache über das Verständnis der einzelnen Items erhöhen die methodische Qualität des vorliegenden Reviews.

Limitationen des Reviews

Die Limitationen in der eigenen Methodik begrenzen die Aussagekraft der Ergebnisse dieses Reviews. Hier ist die Möglichkeit eines Sprachbias aufgrund der Beschränkung auf englische und deutsche Literatur zu nennen.

Außerdem wurden lediglich 4 Studien eingeschlossen, was eine zuverlässige Beurteilung der Reliabilität aufgrund der wenigen Ergebnisse besonders hinsichtlich der vielen unterschiedlichen und oftmals nur einmal untersuchten Tests erschwert. Die geringe Anzahl von nur 4 identifizierten relevanten Studien könnte zum einen an dem erst in den letzten Jahren zunehmenden physiotherapeutischen Forschungsinteresse zum Thema Bewegungskontrolle bei Patienten mit Nackenschmerzen liegen. Dies wird bei der Betrachtung der hohen Aktualität der Studien deutlich. Zum anderen ließe sich die geringe Studienzahl auch mit einem möglichen Publikationsbias erklären. Möglicherweise wurden Studien mit negativen Ergebnissen bezüglich der Reliabilität oder Validität von Bewegungskontrolltests gar nicht publiziert, was zu einer Verzerrung der Ergebnisse geführt haben könnte.

Die dargestellten Ergebnisse könnten zudem unvollständig sein, weil die technikbasierten Tests nicht eingeschlossen wurden. Einige der eingeschlossenen Studien untersuchten auch solche Tests. Die lückenhafte Darstellung könnte unter Umständen hochgradig zuverlässige Tests missachtet haben. Da jedoch das Kriterium der Anwendbarkeit im physiotherapeutischen Praxisalltag von höchster Relevanz ist, lässt sich der Ausschluss von Tests mit technischen Geräten damit begründen.

Ferner ist die verwendete QAREL-Checkliste zur Beurteilung der methodischen Studienqualität bisher nicht auf Deutsch validiert.

Schlussfolgerungen

Der vorliegende systematische Review untersuchte als erster die Intra- und Intertester-Reliabilität von Bewegungskontrolltests der HWS ohne technische Geräte. Zuverlässige Messinstrumente bzw. Testverfahren zur Untersuchung von Patienten mit HWS-Beschwerden sind eine notwendige Voraussetzung für eine bedarfsgerechte physiotherapeutische Versorgung. Die Zahl von Patienten mit unspezifischen Nackenschmerzen ist hoch. Testverfahren müssen eine Klassifizierung der Patienten in Subgruppen ermöglichen. Eine dieser Subgruppen beinhaltet Patienten mit Bewegungskontrollproblemen. Geeignete reliable und valide Tests sollten Patienten dieser Subgruppe identifizieren können.

Neben der Möglichkeit der Klassifizierung von Patienten sind Testverfahren zur Erstellung eines Wiederbefundes, zur Verlaufskontrolle und zur Evaluation einer Behandlung von Bedeutung. Hierbei ist insbesondere für den physiotherapeutischen Praxisalltag wichtig, dass derselbe Therapeut bei wiederholten Messungen unter gleichen Bedingungen ebenso wie verschiedene Therapeuten zu gleichen Testergebnissen kommen.

Die Kriteriumsvalidität der verschiedenen Tests wurde bislang nicht untersucht. Dies liegt besonders an dem bisher nicht vorhandenen Goldstandard zur Beurteilung der Bewegungskontrolle von Patienten mit Nackenschmerzen. Hier ist weitere Forschung besonders bei den als reliabel bewerteten Tests notwendig.

Insgesamt kann tendenziell von einer moderaten bis sehr guten Intratester-Reliabilität und einer ausreichenden bis sehr guten Intertester-Reliabilität der untersuchten Tests ausgegangen werden. Während die Intertester-Reliabilität bei 25 von 31 Testungen (26 verschiedenen Tests, von denen 3 in 2 Studien und 2 Tests in 1 Studie zu unterschiedlichen Messzeitpunkten untersucht wurden) gute bis sehr gute Werte erreichte, erreichte die Intratester-Reliabilität der insgesamt 11 Tests (von jeweils 2 Untersuchern getestet) 21-mal eine gut oder sehr gute Bewertung. Diese Beurteilung der Tests ist aufgrund der großen Spannweite der 95 %-Konfidenzintervalle der einzelnen Kappa-Werte mit Vorsicht zu betrachten.

Nach dem aktuellen Stand der Ergebnisse empfehlen die Autorinnen, nur die Tests mit einer guten bis sehr guten Reliabilität einzusetzen. Tendenziell am ehesten zu empfehlen ist der Test „Blickstabilität“, der als einer der wenigen Tests von 2 Studien untersucht und als reliabel bewertet wurde [11, 14].

Zukünftige Studien sollten neben der Untersuchung der Validität die sehr reliablen Tests zu einer Testbatterie kombinieren und mit einzelnen Tests vergleichen. Sie sollten außerdem die verschiedenen Tests untersuchen und dabei eine einheitliche Methodik, z. B. im Hinblick auf die Definition der Patientenpopulation und Datenauswertung anwenden.

Interessenkonflikt

Die Autoren geben an, dass kein Interessenkonflikt besteht.

Literatur

[1] Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Med 1998; 26: 217 – 238

[2] Bossuyt PM, Reitsma JB, Bruns DE et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Standards for Reporting of Diagnostic Accuracy. Clin Chem 2003; 49: 1 – 6

[3] Boswell MV, Manchikanti L, Kaye AD et al. A Best-Evidence Systematic Appraisal of the Diagnostic Accuracy and Utility of Facet (Zygapophysial) Joint Injections in Chronic Spinal Pain. Pain Physician 2015; 18: E497 – E533

[4] Carlsson H, Rasmussen-Barr E. Clinical screening tests for assessing movement control in non-specific low-back pain. A systematic review of intra- and inter-observer reliability studies. Man Ther 2013; 18: 103 – 110

[5] Childs JD, Fritz JM, Piva SR et al. Proposal of a classification system for patients with neck pain. J Orthop Sports Phys Ther 2004; 34: 686 – 696; discussion: 697–700

[6] Childs JD, Cleland JA, Elliott JM et al. Neck pain: Clinical practice guidelines linked to the International Classification of Functioning, Disability, and Health from the Orthopedic Section of the American Physical Therapy Association. J Orthop Sports Phys Ther 2008; 38: A1 – A34

[7] Comerford M, Mottram S. Kinetic Control – The Management of Uncontrolled Movement. St. Louis: Elsevier; 2012

[8] Cote P, van der Velde G, Cassidy JD et al. The burden and determinants of neck pain in workers: results of the Bone and Joint Decade 2000–2010 Task Force on Neck Pain and Its Associated Disorders. Spine 2008; 33: S60 – S74

[9] Dankaerts W, O'Sullivan PB, Straker LM et al. The inter-examiner reliability of a classification method for non-specific chronic low back pain patients with motor control impairment. Man Ther 2006; 11: 28 – 39

[10] Dankaerts W, OSullivan P. The validity of O'Sullivans classification system (CS) for a sub-group of NS-CLBP with motor control impairment (MCI): overview of a series of studies and review of the literature. Man Ther 2011; 16: 9 – 14

[11] Della Casa E, Affolter Helbling J, Meichtry A et al. Head-eye movement control tests in patients with chronic neck pain; inter-observer reliability and discriminative validity. BMC Musculoskelet Disord 2014; 15: 16

[12] Hogg-Johnson S, van der Velde G, Carroll LJ et al. The burden and determinants of neck pain in the general population: results of the Bone and Joint Decade 2000–2010 Task Force on Neck Pain and Its Associated Disorders. Spine 2008; 33: S39 – S51

[13] International Association for the Study of Pain (IASP). Neck Pain. 2009 www.iasp-pain.org/files/Content/ContentFolders/GlobalYearAgainstPain2/MusculoskeletalPainFactSheets/NeckPain_Final.pdf. (31.05.2017)

[14] Jorgensen R, Ris I, Falla D et al. Reliability, construct and discriminative validity of clinical testing in subjects with and without chronic neck pain. BMC Musculoskelet Disord 2014; 15: 408

[15] Jull G, Falla D, OLeary S et al. Cervical Spine: Idiopathic Neck Pain. In: Jull G, Moore A, Falla D, et al., (eds) Grieves Modern Musculoskeletal Physiotherapie. St. Louis: Elsevier; 2015

[16] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33: 159 – 174

[17] Lucas NP, Macaskill P, Irwig L et al. The development of a quality appraisal tool for studies of diagnostic reliability (QAREL). J Clin Epidemiol 2010; 63: 854 – 861

[18] Luomajoki H, Kool J, de Bruin ED et al. Reliability of movement control tests in the lumbar spine. BMC Musculoskelet Disord 2007; 8: 90

[19] Makela M, Heliovaara M, Sievers K et al. Prevalence, determinants, and consequences of chronic neck pain in Finland. Am J Epidemiol 1991; 134: 1356 – 1367

[20] Michiels S, De Hertogh W, Truijen S et al. The assessment of cervical sensory motor control: a systematic review focusing on measuring methods and their clinimetric characteristics. Gait Posture 2013; 38: 1 – 7

[21] Niere KR, Torney SK. Clinicians' perceptions of minor cervical instability. Man Ther 2004; 9: 144 – 150

[22] O'Leary S, Falla D, Elliott JM et al. Muscle dysfunction in cervical spine pain: implications for assessment and management. J Orthop Sports Phys Ther 2009; 39: 324 – 333

[23] Oesch P, Hilfiker R, Keller S et al. Assessments in der Rehabilitation. Bd. 2: Bewegungsapparat Bern: Huber; 2011

[24] O'Sullivan P. Diagnosis and classification of chronic low back pain disorders: maladaptive movement and motor control impairments as underlying mechanism. Man Ther 2005; 10: 242 – 255

[25] Patroncini M, Hannig S, Meichtry A et al. Reliability of movement control tests on the cervical spine. BMC Musculoskelet Disord 2014; 15: 402

[26] Van Randeraad-van der Zee CH, Beurskens AJ, Swinkels RA et al. The burden of neck pain: its meaning for persons with neck pain and healthcare providers, explored by concept mapping. Qual Life Res 2016; 25: 1219 – 1225

[27] Sahrmann S. Movement System Impairment Syndromes of Extremitites, Cervical and Thoracic Spines. St. Louis: Elsevier; 2011

[28] Scherer M, Plat E. Nackenschmerzen DEGAM-Leitlinie Nr.13. Düsseldorf: Deutsche Gesellschaft für Allgemeinmedizin und Familienmedizin. 2009

[29] Schneider E, Irastorza X, Copsey S. OSH in figures: Work-related musculoskeletal disorders in the EU – Facts and figures. Düsseldorf: Dictus; 2011

[30] Segarra V, Duenas L, Torres R et al. Inter-and intra-tester reliability of a battery of cervical movement control dysfunction tests. Man Ther 2015; 20: 570 – 579

[31] Shumway-Cook A, Woollacott MH. Motor Control: Translating Research Into Clinical Practice. Philadelphia: Lippincott Raven; 2011

[32] Simopoulos TT, Manchikanti L, Gupta S et al. Systematic Review of the Diagnostic Accuracy and Therapeutic Effectiveness of Sacroiliac Joint Interventions. Pain Physician 2015; 18: E713 – E756

[33] Statistisches Bundesamt. Gesundheit. Krankheitskosten 2002, 2004, 2006 und 2008. 2015 www.destatis.de/DE/Publikationen/Thematisch/Gesundheit/Krankheitskosten/Krankheitskosten2120720159004.pdf;jsessionid=B9DAA99AD77875354189C962544A85C2.InternetLive2?__blob=publicationFile (31.05.2017)

[34] Terwee C. Protocol for systematic reviews of measurement properties. 2011 www.cosmin.nl/images/upload/files/Protocol%20klinimetrische%20review%20version%20nov%202011.pdf (31.05.2017)

[35] The Pain Proposal Steering Committee. Pain Proposal – Improving the current and future management of chronic pain. 2010 www.dgss.org/fileadmin/pdf/Pain_Proposal_European_Consensus_Report.pdf (31.05.2017)

[36] Vibe Fersum K, O'Sullivan PB, Kvale A et al. Inter-examiner reliability of a classification system for patients with non-specific low back pain. Man Ther 2009; 14: 555 – 561

[37] Wang WT, Olson SL, Campbell AH et al. Effectiveness of physical therapy for patients with neck pain: an individualized approach using a clinical decision-making algorithm. Am J Phys Med Rehabil 2003; 82: 203 – 218; quiz: 219–221

[38] Werneke M, Hart DL, Cook D. A descriptive study of the centralization phenomenon. A prospective analysis. Spine 1999; 24: 676 – 683

[39] Whiting P, Rutjes AW, Reitsma JB et al. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol 2003; 3: 25

[40] Whiting PF, Rutjes AW, Westwood ME et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011; 155: 529 – 536