



UNIVERSITÀ DEGLI STUDI
DI GENOVA



campus savona

Università degli Studi di Genova

Scuola di Scienze Mediche e Farmaceutiche

Dipartimento di Neuroscienze, Riabilitazione, Oftalmologia, Genetica e Scienze Materno-
Infantili

Master in Riabilitazione dei Disordini Muscoloscheletrici

A.A 2017/2018

Campus Universitario di Savona

**Accuratezza e affidabilità delle procedure di palpazione
vertebrale volte all'identificazione del livello ipomobile.**

Revisione della letteratura

Candidato:

Dott. Ft. Roberta Bellini

Relatore:

dott. Ft. OMT Erica Barboni

INDICE

ABSTRACT

1 – INTRODUZIONE

1. Background
2. Reliability
3. Accuratezza
4. Obiettivo dello studio

2 – MATERIALI E METODI

1. Fonti dei dati
2. Criteri di inclusione ed esclusione
3. Selezione degli studi
4. Outcome degli studi
5. Valutazione del rischio di bias

3 – RISULTATI

1. Qualità metodologica degli studi
2. Riepilogo dei risultati
3. Rachide lombare
4. Rachide dorsale
5. Rachide cervicale

4 – DISCUSSIONE

5 – CONCLUSIONI

BIBLIOGRAFIA

ABSTRACT

Background: Le tecniche di palpazione vertebrale sono uno degli strumenti da sempre utilizzati dai fisioterapisti per identificare la mobilità segmentale passiva. Ciò nonostante, al momento in letteratura si evidenzia come vi sia una scarsa concordanza e una ridotta evidenza sull'accuratezza e l'affidabilità intra- e inter-esaminatore dei test di motion palpation. Inoltre non vi è chiarezza su quale sia la tecnica più accurata ed affidabile per identificare le alterazioni di movimento dei vari segmenti vertebrali.

Obiettivo: L'obiettivo è effettuare una revisione sistematica degli articoli presenti in letteratura sulla reale accuratezza e affidabilità intra- e inter-esaminatore delle tecniche di palpazione segmentaria volte ad identificare il livello ipomobile della colonna vertebrale.

Materiali e metodi: Gli studi sono stati selezionati mediante una ricerca bibliografica degli articoli pubblicati in lingua inglese nelle banche dati elettroniche usando come database MEDLINE e PEDRO. Ulteriori studi sono stati individuati attraverso il check delle voci bibliografiche e dalla ricerca manuale su riviste specialistiche. A partire dagli articoli individuati è stato effettuato uno screening manuale dei potenziali articoli eleggibili visionando il titolo e l'abstract, per procedere, in seguito, con la lettura del full text durante il processo di selezione. Sono stati quindi esclusi gli articoli che non rispondevano ai criteri di inclusione. La valutazione del rischio di bias è stata eseguita con la QAREL Checklist.

Risultati: Sono stati selezionati 19 studi di cui 6 ad alto rischio di bias, 5 moderato e 8 a basso rischio. La maggior parte degli studi ha analizzato l'affidabilità riportando valori da lievi a moderati intraoperatore e da sufficienti a moderati interoperatore. Valori maggiori sono stati ottenuti considerando l'accordo ampliato, tuttavia i dati risultano spesso contrastanti e di difficile confronto a causa delle diverse tecniche, degli esaminatori e dei soggetti coinvolti. Lo studio dell'accuratezza ha interessato un numero minore di pubblicazioni e ne ha evidenziato bassi valori.

Conclusioni: Studi di alta qualità per i test di mobilità segmentale sono molto rari, tuttavia, nonostante manchi sicurezza su ciò che viene stimato esattamente, i risultati sull'affidabilità

sono fondamentali per i terapeuti manuali, soprattutto se uniti a quelli ottenuti con la storia clinica del paziente e con il trattamento. Nonostante i dati di accuratezza e affidabilità siano poco incoraggianti, tali tecniche non devono essere eliminate nella valutazione della mobilità del rachide, dato che nella pratica clinica l'esame obiettivo comprende più tecniche e le conclusioni non sono mai basate solo sui risultati di un singolo test.

Parole chiave: Palpation, motion palpation, mobility, spinal, spinal diseases, thoracic spine, lumbar spine, neck, validity, sensitivity and specificity, accuracy, reliability, reproducibility, test-retest, inter-examiner and intra-examiner, inter-rater and intra-rater.

1 – INTRODUZIONE

1.1 Background

Le tecniche di palpazione vertebrale sono uno degli strumenti da sempre utilizzati dai fisioterapisti nella loro pratica clinica per identificare i segmenti vertebrali. Localizzati i segmenti, la valutazione del movimento intervertebrale è considerata una componente fondamentale dell'esame clinico da parte di molti terapisti manuali con l'obiettivo di individuare una mobilità ristretta, alterata o anomala. Una ricerca effettuata sui fisioterapisti australiani ha evidenziato che la maggioranza (98%) dei fisioterapisti manuali utilizza valutazioni della mobilità spinale durante il loro esame obiettivo e che le decisioni di trattamento si basano sui risultati ottenuti. (1)

Secondo quanto riportato nella revisione di Haas et al., quattro ipotesi formano la logica per l'uso delle tecniche di palpazione del movimento come strumento diagnostico:

1. le anomalie del movimento segmentale a livello spinale causano o contribuiscono alla limitazione funzionale e alla disabilità,
2. la palpazione del movimento è un indicatore affidabile di queste anomalie del movimento,
3. la palpazione del movimento è un indicatore valido di queste anomalie,
4. la palpazione del movimento è sensibile ai cambiamenti clinicamente importanti di queste anomalie del movimento. (2)

Le informazioni ricavate da queste tecniche permettono di integrare quelle rilevate dall'anamnesi e dal resto della valutazione funzionale e aiutano a definire la diagnosi funzionale del paziente ed il programma di trattamento. Alcune delle tecniche utilizzate per la valutazione possono diventare un ponte per la terapia quando utilizzate nel trattamento dell'ipomobilità secondo parametri di forza, durata e frequenza che vanno settati sul singolo paziente.

Tra le varie tecniche, la mobilizzazione posteriore-anteriore (PA) come descritto da Maitland et al. (2005), è una delle più comunemente utilizzate. Il dosaggio è guidato da un sistema di classificazione che si basa sulla percezione del terapeuta della rigidità della vertebra trattata,

che può arrivare all'applicazione di una manipolazione spinale qualora reputi che possa portare beneficio.

Poiché i metodi di valutazione del movimento passivo segmentale sono insegnati ai fisioterapisti fin dall'inizio dei loro percorsi di studi e, come detto prima, vengono utilizzati nella pratica clinica è importante esaminare se questi test sono affidabili. (3-4)

Al momento non vi è chiarezza su quale sia la tecnica più accurata, ripetibile ed affidabile per identificare le alterazioni di movimento dei vari segmenti vertebrali. In letteratura si evidenzia come vi sia una scarsa concordanza e una ridotta evidenza dell'accuratezza e affidabilità intra- e inter-esaminatore di tali tecniche di palpazione della mobilità segmentale.

Molti studi sono stati condotti indagando l'affidabilità di questi test passivi con risultati variabili, in parte anche a causa del fatto che le definizioni delle tecniche di valutazione non erano chiare. Un esempio è lo studio di Pool et al., 2004 (5), in cui è stata descritta solo la direzione della tecnica utilizzata per valutare uno specifico segmento, ad esempio una tecnica in flessione per quanto riguarda l'articolazione atlo-occipitale e in rotazione per l'articolazione atlo-assiale, ma mancava una vera e propria standardizzazione della procedura.

Inoltre, molti studi precedenti si sono concentrati molto sulla provocazione del dolore durante le tecniche di mobilità segmentale, indagando sia soggetti sintomatici che asintomatici.

Negli ultimi anni sono stati sviluppati anche vari dispositivi meccanici in grado di replicare le tecniche di palpazione manuale utilizzando una forza standardizzata per quantificare la resistenza del tessuto alla forza erogata. Questi dispositivi hanno migliorato la capacità di misurare la rigidità spinale, consentendo anche un'indagine più obiettiva sul rapporto tra rigidità articolare spinale e dolore. (6-7)

1.2 Reliability

L'affidabilità è uno dei parametri importanti per determinare la qualità di uno strumento o di una tecnica. Questa misura indica la precisione con cui i risultati dei test possono essere replicati in condizioni identiche o l'assenza di errori di misurazione. Sono stati utilizzati vari

metodi statistici per testare l'affidabilità degli strumenti medici con esiti quantitativi o continui, tuttavia molti di essi, come i limiti di accordo di Bland Altman, si sono dimostrati inappropriati per valutarla. (8)

Tra i vari metodi statistici utilizzati, il coefficiente di correlazione Intraclass (ICC) proposto da Ronald Fisher è quello più popolare per misurare l'affidabilità di variabili continue, mentre per le variabili ordinali viene utilizzato l'indice di concordanza di Cohen (k).

L'affidabilità può essere suddivisa in due tipi: inter- ed intra-esaminatore. La prima valuta il grado con cui due o più esaminatori danno stime consistenti dello stesso fenomeno. L'affidabilità intra-esaminatore valuta, invece, la capacità di un esaminatore di ottenere risposte ripetibili nel tempo, in assenza di cambiamenti, attraverso due o più prove.

L'affidabilità della valutazione del movimento intervertebrale è stata ampiamente studiata e riesaminata sistematicamente per ciascuno dei livelli della colonna vertebrale, tuttavia le stime di affidabilità di tale movimento mostrano notevoli carenze metodologiche nella maggior parte di questi studi. Dalla letteratura risulta più facile ottenere una buona affidabilità intra-esaminatore rispetto all'affidabilità inter-esaminatore, in quanto quest'ultima richiede un'elevata standardizzazione dei metodi e dell'accordo clinico.

1.3 Accuratezza

Per accuratezza di una misurazione si intende la differenza tra la reale misurazione e il valore della misurazione ottenuta da una fonte esterna attendibile (gold standard), o la percentuale di differenza tra i due valori.

In questo contesto si intende la capacità di identificare correttamente un livello vertebrale preciso e la sua mobilità.

1.4 Obiettivo dello studio

L'obiettivo è effettuare una revisione sistematica degli articoli presenti in letteratura sulla reale accuratezza e affidabilità intra- e inter-esaminatore delle tecniche di palpazione segmentaria volte ad identificare il livello ipomobile della colonna vertebrale.

2 - MATERIALI E METODI

2.1 Fonti dei dati

Gli studi sono stati selezionati mediante una ricerca bibliografica degli articoli pubblicati in lingua inglese nelle banche dati elettroniche usando come database MEDLINE e PEDRO. La ricerca è stata condotta fino a Marzo 2019, senza quindi adottare limiti temporali.

Per quanto riguarda la banca dati MEDLINE si è utilizzata l'interfaccia Pubmed e la stringa utilizzata è stata:

("Palpation"[Mesh] OR motion-palpat* OR flexion-extension-test* OR spring-test* OR palpation-test* OR palpatory-diagnos* OR palpatory-test*) AND ("Spinal Diseases"[Mesh] OR "Back Pain"[Mesh] OR "Neck Pain"[Mesh] OR "Spine"[Mesh] OR "Neck"[Mesh] OR spine OR spinal OR vertebra* OR column OR lumbar OR thoracic OR back OR neck OR cervica*) AND ("Reproducibility of Results"[Mesh] OR "Sensitivity and Specificity"[Mesh] OR reproducibil* OR sensitiv* OR specificit* OR reliab* OR valid* OR accur* OR test-retest* OR interrater* OR inter-rater* OR interexamin* OR inter-examin* OR interobserv* OR inter-observ* OR intrarater* OR intra-rater* OR intraexamin* OR intra-examin* OR intraobserv* OR intra-observ*).

Per quanto riguarda la banca dati PEDRO si è condotta una ricerca mediante la seguente stringa di ricerca: Advanced search/Abstract & Title: palpat*.

Ulteriori studi sono stati individuati attraverso il check delle voci bibliografiche e dalla ricerca manuale su riviste specialistiche come: *Archives of Physical Medicine and Rehabilitation*, *Manual Therapy*, *Musculoskeletal Science and Practice*, *The Journal of Manual & Manipulative Therapy* e *Journal of Manipulative and Physiological Therapeutics*.

A partire dagli articoli individuati è stato effettuato uno screening manuale dei potenziali articoli elegibili visionando il titolo e l'abstract, per procedere, in seguito, con la lettura del full text durante il processo di selezione.

I dati estratti riguardavano i soggetti dello studio (numero, età, sesso, condizione clinica), i valutatori (numero, professione, training, esperienza), i metodi di valutazione (postura del paziente, livello vertebrale e tipologia di ROM analizzato, direzione di movimento e spiegazione della tecnica palpatoria) e i risultati (dati di affidabilità intra ed inter esaminatore). Sono stati quindi esclusi gli articoli che non rispondevano ai criteri di inclusione.

La revisione è stata scritta seguendo le indicazioni del Preferred Reporting Items for Systematic Reviews and Meta-analyses: PRISMA.

2.2 Criteri di inclusione ed esclusione

Al fine di essere compresi nella revisione, gli studi presi in esame dovevano presentare un disegno inter-rater, intra-rater o test-retest reliability, avere come distretto d'esame il rachide in toto e avere come obiettivo l'identificazione del segmento ipomobile tramite tecniche di palpazione segmentaria. Gli studi dovevano prevedere indici di misurazione statistica dell'affidabilità (ICC o K).

Si è deciso di escludere gli articoli che non analizzassero la reliability o l'accuratezza ma altri domini o che non riportassero test statistici adeguati, quelli il cui scopo primario non coincidesse con l'outcome prefissato o che avessero come scopo quello di identificare patologie strutturali del rachide e quelli scritti in una lingua che non fosse l'italiano o l'inglese.

Sono stati esclusi gli articoli di cui non è stato possibile reperire il full text.

2.3 Selezione degli studi

La stringa di ricerca ha prodotto al 12/03/2019 490 risultati e per garantire una completezza della ricerca è stato effettuato uno screening manuale dei potenziali articoli elegibili visionando il titolo e l'abstract. L'operazione ha condotto ad un numero di 14 articoli inclusi nel processo di elegibilità di cui è stata effettuata un'analisi del full text.

Per quanto riguarda la banca dati PEDRO si è condotta una ricerca mediante la seguente stringa di ricerca: Advanced search/Abstract & Title: palpat*.

Sono stati individuati 121 risultati al 12/03/2019 di cui potenzialmente rilevanti 4, che, dopo la lettura degli abstract, si è visto non rispettare i criteri di inclusione.

Dal *Journal of Manipulative and Physiological Therapeutics* inserendo come chiave di ricerca “motion palpation” sono stati individuati 427 articoli di cui 11 rilevanti che, eliminati i doppi, si sono ridotti a due articoli che rispettavano i criteri di inclusione.

Ulteriori 10 articoli sono stati ricavati attraverso il check delle voci bibliografiche di articoli ottenuti dalla stringa di ricerca su Pubmed.

Si è svolta la lettura del full text di 25 articoli, di questi cinque sono stati esclusi perchè non rispettavano i criteri di inclusione. Sono stati analizzati, quindi, un totale di 19 articoli inclusi nella sintesi qualitativa.

Il processo di screening è stato schematizzato nel diagramma di flusso creato secondo le linee guida del PRISMA Statement 2015. (Fig. 1)

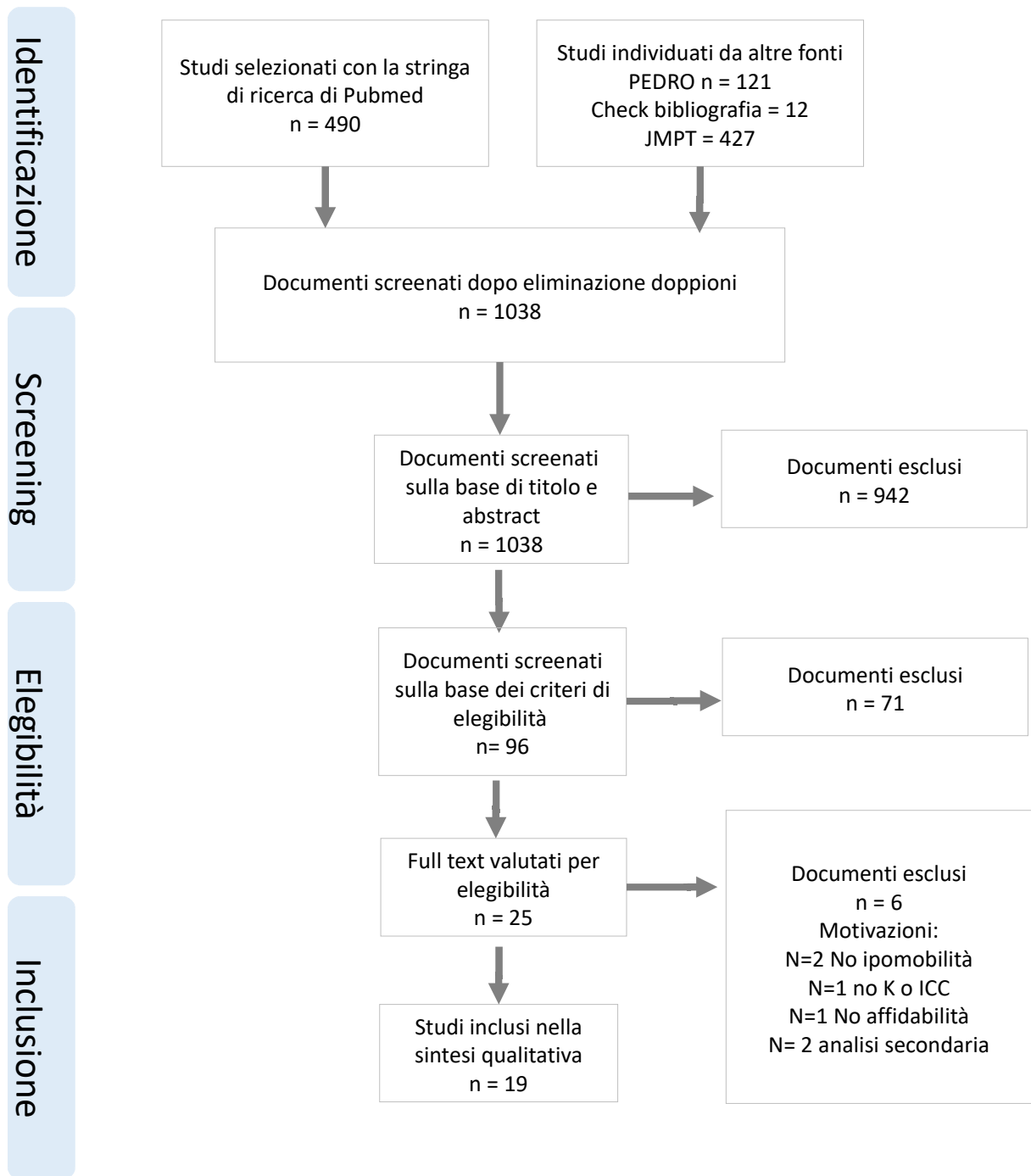


Fig.1 : From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009).
Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097.
doi:10.1371/journal.pmed1000097

2.4 Outcome analizzati

L'affidabilità è stata valutata mediante l'indice di concordanza di Cohen (k), concordanza al netto del caso, per quanto riguarda le variabili ordinali e nominali. I valori di Kappa esprimono di quanto l'accordo supera quello che ci si sarebbe aspettato per caso, mentre il Kappa pesato indica di quanto i risultati deviano tra loro stessi, ponderando il disaccordo in funzione dell'entità della discrepanza. Tenzialmente la Kappa ponderata è maggiore di quella non ponderata in quanto i disaccordi sono più probabili in una sola categoria che in diverse.

I valori vengono interpretati con scala di valutazione di Landis e Koch nella maggior parte degli studi e variano da -1 (massimo disaccordo) ad 1 (accordo perfetto). Valori < 0.10 non indicano alcun accordo, valori da 0.11 a 0.20 lieve accordo, da 0.21 a 0.40 accordo sufficiente, da 0.41 a 0.60 accordo moderato, da 0.61 a 0.80 accordo sostanziale e valori > 0.81 indicano accordo quasi perfetto o eccellente. Il cut-off per considerare una misurazione concordante è pari allo 0.60.

Le variabili qualitative, invece, sono state analizzate mediante il coefficiente di correlazione intra-classe (ICC). Anche in questo caso i valori sono compresi tra 0 (nessuna correlazione) e 1 (massima correlazione), con valori interpretati come segue: < 0.10 nessuna; 0.11-0.40 lieve; 0.41-0.60 discreta; 0.61-0.80 moderato; e valori > 0.81 sostanziale affidabilità.

2.5 Valutazione del rischio di bias

La valutazione del rischio di bias degli articoli inclusi nella revisione è stata eseguita utilizzando la scala di valutazione metodologica Quality Appraisal of Reliability Studies (QAREL).

Si tratta di un recente strumento usato per gli studi di affidabilità diagnostica che ha messo in luce buoni valori di affidabilità (9).

Ogni criterio è stato classificato come Sì (Y), No (N) e Incerto (U) nel caso in cui non vi fossero sufficienti informazioni e Non Applicabile (N/A). I criteri giudicati come Incerti o No sono stati giudicati ad alto rischio di bias.

La checklist include 11 items e il rischio di bias complessivo di ogni studio è stato calcolato sommando il punteggio ottenuto da ciascun item positivo della checklist. Punteggi < 7 sono

stati considerati ad alto rischio di bias, punteggi = 7 hanno evidenziato un moderato rischio di bias e punteggi > 8 sono stati valutati a basso rischio di bias.

La checklist è stata riportata nell'immagine sottostante.

Quality Appraisal of Diagnostic Reliability (QAREL) Checklist

Item	Yes	No	Unclear	N/A
1. Was the test evaluated in a sample of subjects who were representative of those to whom the authors intended the results to be applied? (DEF: 3, 4, 5, 7, 8, 9)	☐	☐	☐	
2. Was the test performed by raters who were representative of those to whom the authors intended the results to be applied? (DEF 3, 4, 6, 7, 8, 9)	☐	☐	☐	
3. Were raters blinded to the findings of other raters during the study? (DEF 10)	☐	☐	☐	☐
4. Were raters blinded to their own prior findings of the test under evaluation? (DEF 11)	☐	☐	☐	☐
5. Were raters blinded to the results of the reference standard for the target disorder (or variable) being evaluated? (DEF 12)	☐	☐	☐	☐
6. Were raters blinded to clinical information that was not intended to be provided as part of the testing procedure or study design? (DEF 13)	☐	☐	☐	☐
7. Were raters blinded to additional cues that were not part of the test? (DEF 14)	☐	☐	☐	
8. Was the order of examination varied? (DEF 15, 16)	☐	☐	☐	☐
9. Was the time interval between repeated measurements compatible with the stability (or theoretical stability) of the variable being measured? (DEF 17)	☐	☐	☐	
10. Was the test applied correctly and interpreted appropriately? (DEF 18)	☐	☐	☐	
11. Were appropriate statistical measures of agreement used? (DEF 19, 20, 21)	☐	☐	☐	
TOTAL				

Fig.2 : Quality Appraisal of Reliability Studies (QAREL) checklist (Lucas et al., 2010)

3 – RISULTATI

3.1 Qualità metodologica degli studi

Dei 19 studi presi in esame 6 hanno presentato alto rischio di bias (Schneider et al. 2008 (11), Piva et al. 2006 (12), Smedmark et al. 2000 (3), Cleland et al. 2006 (13), Snodgrass et al. 2008 (14), Hanney et al. 2014 (15)), 5 studi hanno mostrato un moderato rischio di bias (Johansson et al. 2006 (16), Landel et al. 2007 (10), Deore et al. 2012 (17), Hicks et al. 2003 (18), Cooperstein et al. 2010 (19)) e 8 studi un basso rischio di bias (Cooperstein et al. 2016 (20), Brisme'e et al. 2006 (21), Heiderscheit et al. 2008 (22), Christensen et al. 2002 (23), Walker et al. 2015 (4), Beynon et al. 2018 (24), Schneider et al. 2013 (25), Cooperstein et al. 2013 (26)).

I principali bias riscontrati riguardano in particolare la mancata chiarezza rispetto alla cecità degli esaminatori riguardo alle patologie dei soggetti o ad altre informazioni cliniche essi riguardanti (items 6 e 7), l'assenza di una randomizzazione nell'ordine delle valutazioni eseguite da operatori indipendenti e di una chiarezza dei tempi intercorsi tra una misurazione e l'altra. In tutti gli studi, ad eccezione dello studio di Landel et al. 2007, è mancato un Reference Standard.

I risultati della checklist per ogni studio incluso sono riportati nella tabella sottostante.

STUDIO	1	2	3	4	5	6	7	8	9	10	11
Johansson 2006	Y	Y	U	N/A	N/A	Y	U	Y	Y	Y	Y
Landel 2007	Y	Y	Y	N/A	Y	U	U	N	Y	Y	Y
Schneider 2008	Y	Y	U	N/A	N/A	U	U	Y	Y	Y	Y
Deore 2012	Y	Y	Y	U	N/A	Y	U	U	Y	Y	Y
Cooperstein 2016	Y	Y	Y	N/A	N/A	Y	U	Y	Y	Y	Y
Hicks 2003	Y	Y	Y	U	N/A	U	U	Y	Y	Y	Y
Brisme'e 2006	Y	Y	Y	N/A	N/A	Y	U	Y	Y	Y	Y
Heiderscheit 2008	Y	Y	Y	U	N/A	Y	U	Y	Y	Y	Y
Cooperstein 2010	Y	Y	Y	N/A	N/A	Y	U	Y	U	Y	Y
Christensen 2002	Y	Y	Y	U	N/A	Y	Y	Y	Y	Y	Y
Walker 2015	Y	Y	Y	N/A	N/A	Y	U	Y	Y	Y	Y
Beynon 2018	Y	Y	Y	Y	N/A	Y	U	Y	U	Y	Y
Piva 2006	Y	Y	N	N/A	N/A	U	U	Y	Y	Y	Y
Schneider 2013	Y	Y	Y	N/A	N/A	Y	Y	Y	Y	Y	Y
Smedmark 2000	Y	Y	U	N/A	N/A	N	U	Y	Y	Y	Y
Snodgrass 2008	Y	Y	N/A	N/A	N/A	U	U	N/A	U	Y	Y
Cleland 2006	Y	Y	Y	N/A	N/A	N	U	N	Y	Y	Y
Cooperstein 2013	Y	Y	Y	N/A	N/A	Y	U	Y	Y	Y	Y
Hannev 2014	Y	Y	Y	N/A	N/A	U	U	N	Y	Y	Y

Tabella 1 : Risultati della QAREL Check-list degli studi inclusi

3.2 Riepilogo dei risultati

Dei diciannove articoli inclusi nella revisione 6 riguardano il livello lombare, 6 quello toracico e 7 il distretto cervicale. Delle pubblicazioni incluse 14 studi hanno analizzato l'affidabilità interesaminatore, 1 studio ha analizzato l'affidabilità intraesaminatore e 4 studi le hanno analizzate entrambe. Parallelamente allo studio dell'affidabilità 15 articoli hanno analizzato anche la percentuale di accordo tra esaminatori, sia come accordo stretto che come accordo espanso. L'affidabilità risulta leggermente migliore quando viene considerato l'accordo espanso in cui è stato permesso agli esaminatori di concordare se si trovassero al medesimo livello spinale.

Un solo studio ha analizzato l'accuratezza (accuracy) delle misurazioni ottenute con l'applicazione di tecniche in PA confrontate con i risultati della risonanza magnetica (gold standard), a differenza della maggior parte degli studi inclusi nella revisione in cui manca un gold standard di riferimento.(10)

Lo studio di Landel et al. ha evidenziato bassi valori di accuratezza (intesa come validity) sia per l'individuazione del segmento più mobile ($k = .00$) che del segmento meno mobile ($k = .04$). I bassi valori di k e delle percentuali di accordo evidenziano come la valutazione manuale della mobilità segmentale lombare non concordava con la misura del movimento segmentale del piano sagittale mediante risonanza magnetica dinamica. Questo risultato mette in discussione ciò che i fisioterapisti percepiscono con tale procedura e conferma che gli studi di validità di alta qualità nei test di mobilità manuale sono molto rari. (16)

Gli studi di Cooperstein et al. oltre ad indagare l'affidabilità interoperatore si concentrano sul misurare la distanza centimetrica dal segmento ritrovato ipomobile e un punto di repere comune. (19, 20, 26) Lo studio che indaga il rachide cervicale analizza anche l'accuratezza di queste misurazioni utilizzando il root mean square error (RMSE) e riporta un livello di accuratezza delle misurazioni pari a 1,22 cm che, supponendo un valore approssimativo di 1,8 cm per livello cervicale, indica che gli esaminatori hanno identificato lo stesso segmento vertebrale come ipomobile.

Gli osservatori impiegati negli studi sono per la maggior parte fisioterapisti (12 studi) di cui 8 specializzati in terapia manuale e in 7 studi vengono impiegati chiropratici con diversi livelli di esperienza clinica.

Nelle tabelle sinottiche riportate di seguito vengono schematizzate le caratteristiche di ogni studio con relativo riferimento bibliografico.

RIFERIMENTO BIBLIOGRAFICO	SOGGETTI	ESAMINATORI	OBIETTIVI	PROCEDURA	RISULTATI
Interexaminer reliability of lumbar segmental mobility tests. Johansson, 2006	20 soggetti tra pazienti con LBP in atto o con un episodio recente e studenti asintomatici.	3 terapisti OMT di 2° livello che non hanno mai lavorato assieme con impossibilità a standardizzare la tecnica.	Investigare l'affidabilità interoperatore dei test di mobilità segmentaria durante la flessione e l'estensione lombare.	Pazienti in decubito laterale con anche e ginocchia flesse e il fisioterapista di fronte al paziente. La mobilità segmentaria è stata valutata palpando il processo spinoso della vertebra target e quello adiacente effettuando una flessione-estensione passiva mediante il movimento delle gambe. Prima della valutazione i soggetti sono stati sottoposti ad un riscaldamento attivo della zona lombare.	Scala di valori in 5 punti da estrema ipomobilità a estrema ipermobilità. K pesato <0.20 in 24 dei 30 confronti e compreso tra 0.21 e 0.40 in 4 confronti. K pesato moderato tra ft2 e ft3 in estensione di L2-L3 e L5-S1. Traslando la scala di valori a 3 punti il k pesato non differiva in maniera sostanziale. Povera affidabilità interoperatore.

RIFERIMENTO BIBLIOGRAFICO	SOGGETTI	ESAMINATORI	OBIETTIVI	PROCEDURA	RISULTATI
Intertester reliability and validity of motion assessments during lumbar spine accessory motion testing. Landel, 2007	29 soggetti con età compresa tra 18 e 45 anni con diagnosi di LBP aspecifico.	2 fisioterapisti con 15 e 16 anni di esperienza. Il primo esegue PA con MRI, mentre il secondo PA senza MRI. Un terzo operatore registra i dati ottenuti dalla MRI.	Determinare l'affidabilità interesaminatore nella valutazione tramite PA della mobilità segmentaria del rachide lombare con scala dicotomica e la validità con MRI dinamica.	La prima misurazione viene effettuata da un fisioterapista con paziente prono e l'utilizzo di MRI. La seconda misurazione viene effettuata da un altro fisioterapista con paziente prono e in assenza di MRI. La posizione delle mani, la forza esercitata e le tempistiche di tenuta della compressione e di rilascio sono state standardizzate.	Buona affidabilità per quanto riguarda il segmento meno mobile con $k=0,71$ e accordo al 82.8%. Povera affidabilità nell'identificare il segmento più mobile con $k=0,29$ nonostante il buon accordo al 79,3%. La validità delle tecniche di PA per identificare il segmento più e meno mobile è altrettanto povera.

RIFERIMENTO BIBLIOGRAFICO	SOGGETTI	ESAMINATORI	OBIETTIVI	PROCEDURA	RISULTATI
Spinal palpation for lumbar segmental mobility and pain provocation: an interexaminer reliability study. Schneider, 2008	39 pazienti (18-65 anni) con storia di LBP che al momento del reclutamento stavano svolgendo trattamenti chiropratici.	Due chiropratici con 25 e 10 anni di esperienza che utilizzano abitualmente tecniche di springing. Sono state svolte due ore di training per standardizzare protocollo e procedure d'esame.	Verificare il livello di affidabilità interesaminatore durante lo springing sulle faccette lombari e le articolazioni sacroiliache per la provocazione del dolore e per la mobilità segmentale. Quest'ultimo è stato anche valutato nel contesto del prone instability test.	Il ricercatore principale marca i reperi ossei principali su tutti i soggetti poi i due esaminatori testano tramite tecniche in PA la mobilità segmentaria (normale o ristretta) e la provocazione del dolore (doloroso o non) sia in posizione prona semplice sia durante il prone instability test.	I dati sono stati valutati con il valore k semplice e il valore k regolato usando gli indici di bias e di prevalenza (PABAK). L'affidabilità per la mobilità segmentaria è povera (k da -.17 a .17), il pain provocation test da discreto a buono (k da .21 a .73) e il Prone instability test ha dimostrato moderata affidabilità: κ da .46 a .54. I dati sulla mobilità sacroiliaca variano molto se considerato il semplice k (da -.10 a -.11) o il k regolato (da .77 a .82).

RIFERIMENTO BIBLIOGRAFICO	SOGGETTI	ESAMINATORI	OBIETTIVI	PROCEDURA	RISULTATI
The inter-rater and intra-rater reliability of passive physiological accessory movement assessment of lumbar spine in novice manual therapists. Deore, 2012	52 soggetti sani (23 maschi e 29 femmine) con età compresa tra i 18-40 anni.	Due fisioterapisti terapeuti manuali con meno di 12 mesi di esperienza clinica.	Verificare l'affidabilità intra e interoperatore dei movimenti accessori fisiologici passivi (PPAMVs) a livello del rachide lombare. Per l'affidabilità intraoperatore si è usato come disegno di studio il test-retest, per quella interoperatore sono stati analizzati i dati ottenuti da due esaminatori.	Un esaminatore marca i reperi ossei per garantire uniformità ai livelli esaminati. La mobilità è stata valutata tramite PA sul processo spinoso da L1 a L5 in cieco da entrambi gli esaminatori. Uno di essi ha ripetuto la valutazione una settimana dopo per analizzare l'affidabilità intraoperatore. Si è utilizzata una scala di valori in 3 punti (ipo-normo-iper).	La percentuale d'accordo <u>interoperatore</u> va da 38% (L4) e 58% (L2) con K tra .00 (L4) e .20 (L1). La percentuale d'accordo <u>intraoperatore</u> tra 38.4% and 57.6%. Kappa tra .24 (L2) e .57 (L1). Povera affidabilità intra e interoperatore.

RIFERIMENTO BIBLIOGRAFICO	SOGGETTI	ESAMINATORI	OBIETTIVI	PROCEDURA	RISULTATI
The reliability of lumbar motion palpation using continuous analysis and confidence ratings: choosing a relevant index of agreement. Cooperstein, 2016	34 studenti universitari chiropratici asintomatici o con dolore alla schiena ≤ 2 su una scala di dolore 0-10.	2 chiropratici con licenza e con 30 e 10 anni di esperienza che utilizzano regolarmente tecniche di motion palpation.	Obiettivo primario è valutare l'affidabilità interesaminatore della palpazione del movimento del rachide lombare individuando la vertebra "più fissata". Obiettivo secondario è esplorare le proprietà di vari indici di accordo tra esaminatori.	I processi spinosi di S2 e T12 sono stati marcati con paziente prono. I soggetti sono stati poi posizionati sul fianco destro con anca e ginocchio sinistro flesso (posizione pre-manipolativa). L'esaminatore ha applicato pressione sul livello testato, pressione AP alle braccia incrociate del soggetto e PA con la sua coscia lungo l'asse del femore per identificare il livello spinale più fissato e il rado di sicurezza. È stata misurata la distanza tra S2 e il livello trovato.	La distanza tra S2 e il segmento più fissato variava da 1,0 a 15,0 cm (media = 6,9 cm). Segmento più fissato: L3. Un esaminatore ha segnato 12/34 (35%) valutazioni come "non fiducioso", l'altro 7/34 (21%). ICC (2, 1) = 0.39 (povero), non correlazione tra confidenza dell'esaminatore e accordo intraoperatore, valori maggiori sono stati ottenuti per il genere maschile, la maggiore distanza VP-S2 e età. Distribuendo i valori con omogeneità ICC è salito a 0,70 (buono). I valori medi variavano da 2,4 a 2,9 cm e MedianAED da 2,1 a 3,2 cm.L'analisi dei limiti di accordo (LOA) di Bland-Altman ha suggerito un alto accordo interesaminatore sul livello più fissato.

RIFERIMENTO BIBLIOGRAFICO	SOGGETTI	ESAMINATORI	OBIETTIVI	PROCEDURA	RISULTATI
Interrater reliability of clinical examination measures for identification of lumbar segmental instability. Hicks, 2003	63 soggetti (38 femmine e 25 uomini) con LBP in atto e assenza di irradiazione sotto al ginocchio, provenienti da altri studi sul LBP e da una clinica ambulatoriale.	4 fisioterapisti con esperienza in campo ortopedico rispettivamente di 2, 6, 8, 4 anni. Sono state create tre coppie (PT1-PT2, PT2-PT3, PT1-PT4) a cui sono stati assegnati compiti diversi.	Valutare l'affidabilità interoperatore dei test diagnostici clinici che sono comunemente usati nella valutazione dei pazienti con sospetta LSI.	Lo stesso protocollo di prova è stato utilizzato per tutti i soggetti indipendentemente dalla fonte di reclutamento. L'esame clinico ha incluso: movimenti aberranti, 2 test speciali per LSI (posterior shear e prone instability tests), la scala di Beighton per la lassità legamentosa (LLS) e la valutazione del movimento intervertebrale da prono tramite PA.	Il k pesato per i test di mobilità è basso per ogni segmento lombare con un range da -.02 a .26. I soggetti sono stati suddivisi come ipermobili o non e come ipomobili o non. L'affidabilità per l'ipermobilità è bassa (k .30) con accordo al 76%. Per l'ipomobilità l'affidabilità è altrettanto bassa (k .18) con accordo al 59%. I valori per la provocazione del dolore sono più alti a ciascun livello spinale con un range da .25 a .55. L'osservazione dei movimenti aberranti, il Beighton LLS e i test per LSI si sono dimostrati più affidabili.

RIFERIMENTO BIBLIOGRAFICO	SOGGETTI	ESAMINATORI	OBIETTIVI	PROCEDURA	RISULTATI
Interrater reliability of a passive physiological intervertebral motion test In the mid-thoracic spine. Brisme'e, 2006	41 studenti universitari (19 maschi e 22 femmine) di età compresa 19-40 anni, sani, assenza di pregressi interventi chirurgici, patologie reumatiche, gravidanza e BMI <30.	Tre fisioterapisti specializzati OMT con almeno 12 anni di esperienza in terapia manuale.	Esaminare l'affidabilità inter esaminatore di un test di mobilità fisiologica intervertebrale passiva di un segmento di movimento medio-toracico della colonna vertebrale.	Il processo spinoso di T6 è stato marcato partendo da C6, usata per la traslazione ventrale durante l'estensione cervicale. I movimenti 3D di T6-T7 (o T5-T6) sono stati valutati con paziente seduto e terapeuta dal lato opposto alla rotazione e sono stati indagati accoppiamenti omolaterali e controlaterali.	La percentuale di accordo varia da 63.4% a 82.5%. I valori di k mostrano un range da 0.27 a 0.65, con un punteggio complessivo pari a 0.41. 22 soggetti hanno mostrato maggiore movimento segmentale con accoppiamento controlaterale, 19 soggetti con accoppiamento omolaterale. Il test ha un'affidabilità debole/moderata quando eseguito su soggetti sani. L'indice di correlazione di Spearman non risulta statisticamente rilevante.

RIFERIMENTO BIBLIOGRAFICO	SOGGETTI	ESAMINATORI	OBIETTIVI	PROCEDURA	RISULTATI
Reliability of joint mobility and pain assessment of the thoracic spine and rib cage in asymptomatic individuals Heiderscheit, 2008	9 soggetti (3 maschi e 6 femmine) asintomatici senza storia di dolore dorso-lombare con età tra 23-35 anni.	Due terapisti manuali con oltre 10 anni di esperienza clinica nelle tecniche di valutazione utilizzate.	Determinare l'affidabilità inter- e intra-operatore della mobilità articolare del rachide dorsale e della gabbia toracica e della valutazione del dolore.	I PAVMs sono stati testati con paziente prono, sono state applicate cinque pressioni in PA sui processi spinosi e trasversali da T1 a T12. La mobilità posteriore delle coste è stata valutata attraverso cinque pressioni in PA a livello degli angoli costali da K1 a K12. I soggetti sono stati posizionati supini per la valutazione della mobilità anteriore costale con cinque pressioni AP applicate alle articolazioni costosternali. La mobilità di ogni articolazione è stata giudicata ipomobile, ipermobile o normale. È stata indagata anche la presenza o assenza di dolore.	<u>Mobilità rachide dorsale:</u> affidabilità intraoperatore tra leggera e discreta ($\kappa=.17-.26$). Affidabilità interoperatore leggera ($\kappa=.15$). Usando l'accordo ampio l'affidabilità intra- è aumentata a buono ($\kappa=.75$ e $.61$), quella inter-moderata ($\kappa=0.59$). Analoga affidabilità per la valutazione della mobilità congiunta della <u>gabbia toracica</u>, con affidabilità intra- discreta ($\kappa=.26$ e $.29$), buona considerando l'ambito dell'accordo ampliato. ($\kappa=.71$ e $.76$). L'affidabilità mobilità dorsale inter- (accordo rigoroso 71%, esteso 85%) inferiore a quella intra- (accordo rigoroso 74-80%, esteso 90-91%).

RIFERIMENTO BIBLIOGRAFICO	SOGGETTI	ESAMINATORI	OBIETTIVI	PROCEDURA	RISULTATI
Interexaminer reliability of thoracic motion palpation using confidence ratings and continuous analysis. Cooperstein, 2010	52 studenti di chiropratica asintomatici (dolore <2 su scala numerica a 11 punti), sono stati esclusi i soggetti che per qualsiasi motivo non riuscivano a tollerare la procedura di palpazione.	Due chiropratici con più di 20 anni di esperienza clinica.	Valutare l'affidabilità interoperatore dei test di MP a livello toracico consentendo all'esaminatore di valutare la propria fiducia nelle sue conclusioni, con metodo analitico di dati continui, per vedere se può influenzare il livello di concordanza.	I soggetti sono stati collocati in posizione prona e sono stati marcati i processi spinosi di T3, T10 e S1. Le tecniche di MP sono state testate tra T3 e T10 ed effettuate con pressione del pollice per valutare la qualità del movimento (joint-Play). Individuato il livello spinale di maggiore fissazione è stato messo un marcatore sulla pelle e l'esaminatore ha indicato se era "molto fiducioso" o "non molto fiducioso" (assenza di un livello significativamente fissato o più di 1 livello). Un assistente ha misurato la distanza dal processo spinoso di S1 al marcatore e ha registrato il grado di fiducia dell'esaminatore.	I soggetti sono stati divisi in sottogruppi. Per il gruppo 1 (52 soggetti) si ha scarso accordo tra operatori con ICC [2,1] = .3110. Per il gruppo 2 (entrambi fiduciosi) l'accordo era buono con ICC [2,1] = .8266. Quando entrambi gli esaminatori non erano fiduciosi si è ottenuto il più basso valore di ICC [2,1] = .3873. Quando ogni esaminatore era "molto fiducioso" i livelli identificati erano molto vicini con una media di 2 cm e 7.10 cm quando nessuno dei due lo era. Ciò corrisponde a un accordo "buono".

RIFERIMENTO BIBLIOGRAFICO	SOGGETTI	ESAMINATORI	OBIETTIVI	PROCEDURA	RISULTATI
Palpation of the upper thoracic spine: an observer reliability study. Christensen, 2002	29 pazienti e 27 controlli sono stati esaminati per l'affidabilità interoperatore e 14 pazienti e 15 controlli per quella intraoperatore.	2 esperti chiropratici.	Valutare l'affidabilità intraosservatore (hour-to-hour e day-to-day) e interosservatore con 3 procedure di palpazione per rilevare una disfunzione della biomeccanica spinale nei primi 8 segmenti del rachide toracico.	Tre tipi di tecniche: (1) la valutazione del joint-play da prono, (2) la valutazione da seduto per lateroflessione e rotazione e (3) la palpazione per la dolorabilità paraspinale. È stato assegnato un valore come "assente" o "presente" per ogni segmento. <u>Intra-</u> : hour-to-hour, i soggetti sono stati riesaminati entro un intervallo di 1,5-2 ore; day-to-day, i soggetti sono stati esaminati alla stessa ora per 2 giorni consecutivi. <u>Inter-</u> : i soggetti sono stati divisi in due metà e esaminati prima da uno poi dall'altro operatore.	<u>Intra-</u> : per la valutazione da seduto e da prono $k = .13$ e $.33$ (hour-to-hour) e $k = .33$ e $.45$ (day-to-day) per l'accordo stretto e $k = .60$ secondo l'accordo ampliato. <u>Inter-</u> : $k = .00$ e $-.03$ per l'accordo stretto e $k = .22$ e $.24$ secondo l'accordo ampliato. Per la dolorabilità paraspinale i valori di K variano tra $.63$ e $.77$ per tutti e 3 i tipi di affidabilità nell'analisi aggregata, indicando una relativamente buona riproducibilità.

RIFERIMENTO BIBLIOGRAFICO	SOGGETTI	ESAMINATORI	OBIETTIVI	PROCEDURA	RISULTATI
Interrater Reliability of Motion Palpation in the Thoracic Spine. Walker, 2015	25 soggetti con età compresa tra 18-70 anni con dolore dorso-lombare in atto o in anamnesi.	2 chiropratici esperti con pregressa esperienza ambulatoriale e docenza accademica.	Indagare l'affidabilità interoperatore delle tecniche di valutazione del movimento della colonna dorsale per percepire restrizioni articolari e dolore. Sono stati utilizzati sia approcci pragmatici che altamente standardizzati.	Sono stati marcati i processi spinosi da T5 a T12. Ogni soggetto è stato esaminato prima con il metodo personale e poi con un metodo standardizzato (Bergman e Peterson) che includeva movimenti passivi di flesso-estensione, lateroflessione e rotazione bilaterale da seduti e test di mobilità intervertebrale segmentale. Si è ricercato il segmento con mobilità ristretto o doloroso.	Per la restrizione segmentale k tra -.27 e .36 a seconda del livello spinale. Per la provocazione del dolore k tra -.38 e .32, valori deludenti. Il punteggio della tecnica pragmatica era peggiore rispetto a quella standardizzata sia per quanto riguarda la mobilità che la provocazione del dolore.

RIFERIMENTO BIBLIOGRAFICO	SOGGETTI	ESAMINATORI	OBIETTIVI	PROCEDURA	RISULTATI
The interrater reliability of static palpation of the thoracic spine for eliciting tenderness and stiffness to test for a manipulable lesion. Beynon, 2018	34 soggetti (20 femmine, 14 maschi) con età da 18 a 37 anni, 17 asintomatici e 17 sintomatici.	2 chiropratici con tre anni di esperienza clinica.	(1) determinare l'accordo interoperatore della palpazione statica del rachide toracico per elicitare la dolorabilità e la rigidità segmentali, determinando l'effetto della formazione standardizzata per gli esaminatori. (2) valutare il consenso degli esperti sul livello di dolorabilità segmentale richiesto per individuare una "lesione manipolabile".	I processi spinosi da T1 a T12 sono stati marcati. 2 valutazioni: la prima con approccio pragmatico (spring-test sulle spinose con paziente prono per ricercare il segmento rigido e valutarne la dolorabilità su scala numerica a 11 punti NPRS), la seconda con approccio standardizzato (Bergmann and Peterson's 2010)	Accordo rigoroso dell'approccio pragmatico per la rigidità spinale k variava da - .31 a .47 (PABAK da - .14 a .71), per l'approccio standardizzato k variava da - .11 a .53 (PABAK da .12 a .76). Accordo rigoroso alla valutazione della dolorabilità segmentale dell'approccio pragmatico K da .22 a .77 e dell'approccio standardizzato K da .25 a .70. Accordo di vertebra espansa per l'approccio pragmatico alla rigidità segmentale k da - .25 a .30 (PABAK da .09 a .89), per quello standardizzato k da - .11 a .30 (PABAK da - .24 a .76).

RIFERIMENTO BIBLIOGRAFICO	SOGGETTI	ESAMINATORI	OBIETTIVI	PROCEDURA	RISULTATI
Inter-tester reliability of passive intervertebral and active movements of the cervical spine. Piva, 2006	30 soggetti (18 femmine, 12 maschi) con dolore al collo/scapola, età tra 18-75 anni e punteggio <60% al Neck Disability Index .	2 fisioterapisti manuali con differenti 10 e 2 anni rispettivamente di esperienza .	(1) determinare l'affidabilità interoperatore dei movimenti passivi intervertebrali (PIM), dell'AROM e gli effetti sulla provocazione dei sintomi (2) stabilire il cambiamento minimo rilevabile (MDC) nell'AROM cervicale (3) determinare l'associazione tra l'AROM e la disabilità.	Valutazione AROM: paziente seduto per flessione-estensione, lateroflessione e rotazione in max flessione e supino per le rotazioni. È stato usato un goniometro. Valutazione PIM: paziente supino, è stata testata l'articolazione atlo-occipitale in lateral glide e lateral displacement, l'atlo-assiale in flessione e lateroflessione, la dolorabilità dei processi trasversi di C1 e C2-C7 con lateral glide.	<u>PIM</u>: valori notevoli per l'ipomobilità articolare atlo-occipitale con lateral glide ($k = .81$) e dolorabilità dei processi trasversali di C1 ($k = .83$). I valori di kappa erano moderati per la riproduzione dei sintomi durante la flessione laterale atlo-assiale ($k = .61$) e a livello di C4 e C6 ($k = .65$ e $.76$). Scarsi valori per la mobilità di C2 ($k = .46$) e la riproduzione dei sintomi durante a C2 e C5 ($k = .42$ e $.55$). Gli altri test bassi valori di k.

RIFERIMENTO BIBLIOGRAFICO	SOGGETTI	ESAMINATORI	OBIETTIVI	PROCEDURA	RISULTATI
Intrarater and interrater reliability of select clinical tests in patients referred for diagnostic facet joint blocks in the cervical spine. Schneider, 2013	56 soggetti con neck pain cronico (37 femmine, 19 maschi) con età compresa tra 21-64 anni e NRS >3.	2 fisioterapisti con 12 e 16 anni di esperienza clinica	Determinare l'affidabilità intra- e inter- di test clinici utilizzati per valutare i pazienti con dolore al collo persistente indirizzati per effettuare il blocco diagnostico delle faccette articolari e determinare una clinical prediction guide per identificare i pazienti più adatti alla procedura.	L'esame è stato eseguito con una sequenza che comprendeva la valutazione della ROM cervicale, ER (extension-rotation), MSE (manual spinal examination), e la palpazione per la dolorabilità segmentale (PST). MSE con paziente prono e applicazione di PA sui processi articolari da C2 a C7. <u>Inter-</u> : i soggetti sono stati valutati dai 2 fisioterapisti in cieco prima del blocco faccettario programmato. <u>Intra-</u> : lo stesso fisioterapista riesaminare lo stesso paziente 7 giorni dopo l'esame iniziale.	Affidabilità intra- e inter- era elevata per il ROM cervicale (ICC da .90 a .97 per le 6 direzioni di movimento). Affidabilità intra- e inter- per la riproduzione dei sintomi durante il ROM (k ponderato da .60 a 1,0). Test ER, k da .72 a .75 (intra-) e da .92 a .93 (inter-). MSE, k da .62 a .88 (intra-) e da .79 a .96 (inter-). PST, k da .51 a .84 (intra-) e da .74 a .96 (inter-). I segmenti in cui era presente maggiore accordo e valori di k maggiori sono stati C4-C5 e C5-C6. Molti soggetti hanno presentato risultati positivi a più di 1 livello spinale.

RIFERIMENTO BIBLIOGRAFICO	SOGGETTI	ESAMINATORI	OBIETTIVI	PROCEDURA	RISULTATI
Inter-examiner reliability in assessing passive intervertebral motion of the cervical spine. Smedmark, 2000	61 pazienti (46 donne, 15 uomini) con età tra 20-71 anni e neck pain aspecifico.	2 fisioterapisti OMT con 25 anni di esperienza clinica di cui 17 di lavoro insieme.	Verificare l'accordo di due fisioterapisti sulla presenza di rigidità segmentale in tre segmenti cervicali e la prima costa in pazienti con problemi alla colonna cervicale.	Sono stati testati 4 test: - rotazione destra C1-C2 - lateroflessione destra C2-C3 - flesso-estensione C7-T1 - springing prima costa a destra. I risultati sono stati classificati come: rigidità (ridotta mobilità) e non rigidità (mobilità normale o aumentata).	Accordo percentuale per tutti i test è stato del 77% ($70 \pm 87\%$). K variava tra .28 e .43, considerato leggero-moderato. C1-C2 (K .28): test positivo su 2 pazienti, negativo in 50 e il mancato accordo in otto pazienti. Un paziente è stato escluso. C2-C3 (K .43): test positivo in 19, negativo in 23 e il mancato accordo in 18 pazienti. Un paziente è stato escluso. C7-T1 (K .30): test positivo in 42, negativo in 6 e non accordo in 13 pazienti. Prima costa: test positivo in 12, negativo in 30 e non accordo in 18.

RIFERIMENTO BIBLIOGRAFICO	SOGGETTI	ESAMINATORI	OBIETTIVI	PROCEDURA	RISULTATI
Measuring the posteroanterior stiffness of the cervical spine. Snodgrass, 2008	67 pazienti asintomatici con età compresa tra 18 e 50 anni e assenza di dolore al collo o mal di testa tale da necessitare di trattamento nei 12 mesi precedenti. 31 soggetti sono tornati per i test ripetuti.	Un fisioterapista con qualifica OMT e una vasta esperienza didattica.	(1) sviluppare uno strumento affidabile e sicuro per misurare la rigidità della colonna vertebrale cervicale (2) documentare la rigidità in un gruppo di individui asintomatici.	(1) l'affidabilità dello strumento è stata testata con misurazioni ripetute utilizzando otto diverse combinazioni di schiuma di densità variabile . (2) I processi spinosi C2 e C7 sono stati marcati e pre-condizionati con 5 PA manuali, come farebbe un clinico per la valutazione. È stato testato prima C7 poi C2. Lo strumento è stato posto a contatto con la pelle del soggetto e la misurazione è stata presa dopo un'espiazione profonda con frequenza oscillatoria impostata a 1Hz.	(1) ICC (2, 1) per le misurazioni ripetute è stata di 0,99 (95% CI 0.93 – 1.00). (2) a C2 il coefficiente di rigidità K è stato calcolato da 7 a 40N, e per C7 da 20 a 70N. La rigidità media a C2 era di 4.58 N/mm (95% CI 4.30 – 4.85), e a C7 era 7.03 N/mm (95% CI 6.50 – 7.57). ICC (2, 1) per misurazioni ripetute = 0,84 (95% CI 0,74 – 0,90). SEM tra le due occasioni C2= 0,53 N/mm e C7= 0.83 N/mm per il C7.

RIFERIMENTO BIBLIOGRAFICO	SOGGETTI	ESAMINATORI	OBIETTIVI	PROCEDURA	RISULTATI
Interrater Reliability of the History and Physical Examination in Patients With Mechanical Neck Pain. Cleland, 2006	22 pazienti con dolore meccanico al collo, età compresa tra 18 e 60 anni e NDI > 10%.	4 fisioterapisti con esperienza clinica da 3 a 23 anni.	Esaminare l'affidabilità interoperatore della storia clinica e dell'esame fisico in pazienti con dolore meccanico al collo.	È stata prima esaminata la storia clinica di ogni paziente con domande specifiche. Poi si è passati all'esame fisico: osservazione della postura, ROM attivo cervicale e toracico, test di forza muscolare, lunghezza muscolare, Chin Tuck Neck Flexion Test (endurance dei flessori profondi) e mobilità spinale. Quest'ultima è stata testata per: C0-C1 con glide AP, C1-C2 in rotazione, C2-C7 e T1-T9 con PA da prono (giudizio: ipo/ipermobile o normale). Ogni segmento è stato valutato anche come dolente o non.	<u>Mobilità vertebrale: k da -.26 a .74, provocazione del dolore durante la valutazione della mobilità k da -.52 a .90.</u>

RIFERIMENTO BIBLIOGRAFICO	SOGGETTI	ESAMINATORI	OBIETTIVI	PROCEDURA	RISULTATI
Interexaminer reliability of cervical motion palpation using continuous measures and rater confidence levels. Cooperstein, 2013	29 studenti chiropratici asintomatici (19 maschi, 10 femmine) con età media di 27,1 anni. Sono stati esclusi i partecipanti con dolore cervicale riferito superiore a 2/10 o intolleranza alla procedura di palpazione.	3 chiropratici, due con più di 20 anni di esperienza clinica e uno con circa 3 anni di esperienza.	Applicare la metodologia di studio della colonna toracica di Cooperstein alla colonna vertebrale cervicale, con analisi continue e valori di confidenza stratificati.	Con paziente prono è stato marcato il processo spinoso di T1. Da supino gli esaminatori hanno testato i livelli da C1 a C7 con sovrappressioni in PA a end range all'aspetto laterale dei pilastri articolari cervicali, creando un'estensione, lateroflessione ipsilaterale e rotazione controlaterale. Dopo aver identificato il livello ipomobile è stato messo un marcatore adesivo nella posizione indicata. L'esaminatore ha detto se era "molto fiducioso" o "non fiducioso" nel ritrovamento dell'ipomobilità. È stata registrata la distanza in centimetri da questo marcatore al segno sul processo spinoso T1 e il rating di fiducia dell'esaminatore.	Calcoli basati su n = 27. Accordo globale sull'ipomobilità è stato del 44%. Esaminatori 1 & 2 k: -.39, p: .09. Esaminatori 1 & 3 k: -.28, p: .17. Esaminatori 2 & 3 k: -.19, p: .39. Per n=27, ICC [2, 1] = 0,61; RMSE indicava che l'accuratezza della misura era di 1,22 cm. MAD = 1,35 cm, IC 95%: 1,12, 1,57 cm. Differenza media interoperatore di circa il 75% di un livello vertebrale.

RIFERIMENTO BIBLIOGRAFICO	SOGGETTI	ESAMINATORI	OBIETTIVI	PROCEDURA	RISULTATI
Inter-rater reliability of select physical examination procedures in patients with neck pain. Hanney, 2014	22 soggetti con età > 18 anni, NP con o senza sintomi unilaterali dell'arto superiore e NDI di 10/50 o superiore.	6 clinici con una media di 9,17 (DS 5,3) anni di esperienza nella terapia fisica ortopedica.	Valutare l'affidabilità interoperatore dell'esame fisico in pazienti con NP meccanico, il secondo esame è stato condotto 24-48 h dopo il primo. L'ipotesi è di un valore più basso di affidabilità quando c'è un periodo di tempo prolungato tra le procedure di esame.	L'esame includeva la valutazione posturale, ROM, test speciali (ULTT A, Spurling e Distraction), la mobilità passiva delle articolazioni, i test di lunghezza muscolare e della forza muscolare. Le valutazioni della mobilità includevano: C0-C1, C1-C2 e il lateral glide del rachide cervicale inferiore. Valutazione: ipo/ipermobile e normale.	<u>Mobilità segmentale:</u> k ponderato da .05 a .52. C4-5 ha dimostrato il più alto livello di affidabilità (k= .43). Cervicale superiore k= .23 e inferiore k= .30. Percentuale d'accordo va da 59.1% per C3-C4 a 70.5% per C4-C5 e CRLF.

3.3 Rachide lombare

Tutti gli studi presi in esame hanno come obiettivo quello di indagare la mobilità segmentaria del rachide lombare, ma in alcuni di essi vengono indagati anche altri aspetti. Nello studio di Hicks et al. l'autore si pone come obiettivo di valutare l'affidabilità interoperatore dei test diagnostici clinici comunemente usati nella valutazione dei pazienti con sospetta LSI (lumbar segmental instability) indagando, oltre alla mobilità segmentale, i movimenti aberranti, la lassità legamentosa e due test speciali per LSI (posterior shear e prone instability tests). (18)

Anche nello studio di Schneider et al. viene indagata la mobilità segmentale sia in posizione prona che durante il prone instability test e, allo stesso tempo, viene valutata anche la provocazione del dolore durante le due procedure. (11)

L'obiettivo secondario dello studio di Cooperstein et al. è stato quello di esplorare le proprietà di vari indici di accordo tra esaminatori tra cui il Bland-Altman Limits of Agreement (LoA), la Mean Absolute Examiner Differences (MeanAED) e la Mediano Absolute Examiner Differences (MedianAED). (20)

L'affidabilità interoperatore è stata testata in tutti e sei gli studi analizzati, mentre solo lo studio di Deore et al. ha analizzato anche quella intraoperatore con un disegno di studio test-retest in cieco a distanza di una settimana. (17)

Per quanto riguarda il tipo di esaminatori impiegati negli studi non si riscontra omogeneità in quanto in quattro studi vengono impiegati fisioterapisti (10, 16, 17, 18) e in due chiropratici (11, 20). Non risulta esserci omogeneità neanche negli anni di esperienza con estremi che vanno da meno di un anno di esperienza clinica nello studio di Deore a 25 e 30 anni di esperienza rispettivamente negli studi di Schneider e Cooperstein.

Rispetto al numero di pazienti analizzati nei vari studi, quattro lavori hanno presentato un gruppo di 20-39 soggetti e due studi hanno esaminato un gruppo di 52 e 63 soggetti. Hicks, Landel e Schneider hanno esaminato popolazioni di pazienti con LBP aspecifico, Johnsson ha studiato un gruppo eterogeneo di 20 soggetti composto da pazienti con LBP in atto o con episodio recente e studenti asintomatici, Cooperstein soggetti asintomatici o con LBP con dolore minore di 2 su una scala a 10 punti. Solo lo studio di Deore prende in esame un gruppo di soggetti sani.

La metodologia d'esame con cui è stata condotta la valutazione della mobilità sui singoli soggetti vede quattro studi (10, 11, 17,18) con paziente prono e applicazione di una forza postero-anteriore e due studi con paziente sul fianco (16, 20). Inoltre, in 4 studi (10, 11, 18, 20) viene utilizzata una procedura standardizzata con eventuale training precedente alla valutazione, in uno studio non viene specificato se la tecnica è stata standardizzata (17) e in uno è esplicitato che non vi è standardizzazione (16).

L'affidabilità interoperatore risulta bassa/povera in cinque studi su sei (Johansson (16): $k < 0.20$ in 24 su 30 misurazioni, tra 0.21 e 0.40 in 4 misurazioni e sopra i 0.40 in 2; Schneider (11): k da -0.17 a 0.17, PABAK da 0.77 a 0.82; Deore (17): k da 0.00 per L4 a 0.20 per L1; Hicks (18): k compreso tra -0.02 e 0.26; Cooperstein (20) : ICC = 0.39, salito a 0.70 dopo una distribuzione omogenea dei risultati).

Nello studio di Landel et al. (10) le valutazioni del movimento vertebrale sono risultate avere una buona affidabilità per valutare il segmento percepito come meno mobile ($k = 0.71$), ma non erano affidabili per valutare il segmento più mobile ($k = 0.29$). Nonostante le differenze nei valori di kappa, le percentuali di accordo tra i 2 esaminatori erano alte e simili per entrambe le valutazioni (82.8% per il segmento ipomobile e 79.3% per quello più mobile). Inoltre si è evidenziato un povero accordo tra i valori di PA e del movimento intervertebrale misurato con MRI ($K = 0.04$, CI = da -0.16 a 0.24 per il segmento meno mobile).

Hicks (18) ha anche analizzato l'affidabilità delle valutazioni su scala dicotomica per la presenza di qualsiasi ipomobilità nella colonna lombare rispetto a nessuna ipomobilità. Anche in questo caso l'affidabilità per una maggiore mobilità era bassa ($k = 0.30$, accordo percentuale 76%) come quella per l'ipomobilità ($k = 0.18$, accordo percentuale 59%).

Cooperstein et al. (20) hanno anche riscontrato valori di MeanAED che andavano da 2,4-2,9 cm e di MedianAED da 2,1 a 3,2 cm durante la misurazione centimetrica tra il segmento giudicato ipomobile e S2. Tutti i valori erano inferiori all'altezza tipica ipotizzata di una vertebra lombare pari a 4,0 cm, suggerendo in media che gli esaminatori concordavano nell'identificare il livello più fissato, o comunque il segmento di movimento che lo conteneva.

L'affidabilità intraoperatore è stata studiata solo nello studio di Deore (17) con k che andava da 0.24 per L2 a 0.57 per L5, anche in questo caso il livello di affidabilità è povero ma risulta comunque maggiore di quella interoperatore.

3.4 Rachide dorsale

La mobilità intervertebrale è stata presa in esame in tutti e sei gli studi, 4 di questi hanno anche analizzato la provocazione del dolore. (4, 22, 23, 24) Tutti gli studi analizzano anche l'accordo tra gli esaminatori con valori molto variabili a seconda del livello segmentale.

Come per il rachide lombare, lo studio di Cooperstein et al. (19) si è posto anche l'obiettivo di valutare la possibile influenza del livello di fiducia dell'esaminatore sull'accordo tra gli esaminatori.

Anche lo studio di Beynon (24) si pone l'obiettivo di valutare il consenso degli esperti sul livello segmentale di dolorabilità richiesto per individuare una lesione manipolabile. In questo studio e in quello di Walker et al. (4) la mobilità vertebrale viene indagata sia con un approccio pragmatico sia con un approccio standardizzato.

Tutti gli studi hanno indagato l'affidabilità interoperatore, mentre solo Heiderscheit (22) e Christensen (23) hanno indagato quella intraoperatore. Il primo ha considerato una rivalutazione dopo quattro ore, mentre il secondo ha utilizzato un disegno hour-to-hour, in cui i soggetti sono stati riesaminati entro un intervallo di 1,5-2 ore, e day-to-day in cui i soggetti sono stati esaminati alla stessa ora per 2 giorni consecutivi.

Gli esaminatori impiegati negli studi vedono diverse figure professionali e relative specializzazioni: quattro studi vedono due chiropratici con diversi anni di esperienza, dai 3 anni di Beynon (24) ai più di 20 di Cooperstein (19), mentre solo gli studi di Brisme'e (21) e Heiderscheit (22) vedono l'utilizzo di fisioterapisti manuali con 12 e 10 anni di esperienza clinica rispettivamente.

I soggetti reclutati negli studi si presentano eterogenei in quanto tre di essi prendono in esame soggetti sani o asintomatici (19, 21, 22), uno studio ha analizzato soggetti con LBP in atto o recente (4), mentre Brisme'e e Christensen hanno analizzato un gruppo composto da soggetti sintomatici e controlli asintomatici. Christensen (23) per studiare l'affidabilità interoperatore ha considerato un gruppo di 29 pazienti e 27 controlli, mentre per l'affidabilità intraoperatore un gruppo di 14 pazienti e 15 controlli. Negli altri studi il numero dei soggetti era compreso tra 9 (22) e 52 (19).

La metodologia con cui sono stati condotti gli studi è varia, alcuni hanno utilizzato una tecnica in postero-anteriorità con paziente prono (22, 19, 23, 24), mentre in altri studi il paziente è stato valutato da seduto (4, 21, 23). Christensen ha valutato il joint play articolare da prono con PA e da seduto per le rotazioni, lateroflessioni e per la dolorabilità paraspinale.

Tre studi hanno confrontato l'affidabilità di due approcci diversi, uno pragmatico (di utilizzo proprio del singolo esaminatore) e uno standardizzato mediante un training tra gli esaminatori pre-valutazione (4, 22, 24).

I risultati della mobilità vertebrale evidenziano una scarsa affidabilità interoperatore, con molti risultati che analizzati secondo un accordo rigoroso risultano non più affidabili del caso ($k < 10$).

Lo studio di Brisme'e et al. (21) riguardante i segmenti medio-toracici T5-T6 e T6-T7 con tecniche in 3D omo e contro-laterali ha mostrato valori di affidabilità da scarsi a moderati ($k = 0.27 - 0.65$) con un accordo interesaminatore che variava da 63,4% a 82,5%.

Lo studio di Hederscheit (22) ha evidenziato scarsi valori di affidabilità con valori di $k = 0.15$ che, applicando un'estensione dell'accordo, salgono a $k = 0.59$ portando l'affidabilità interesaminatore

da lieve a moderata. L'accordo interoperatore va da 74-80% a 90-91%. Valori simili sono stati individuati negli studi di Christensen (23) e Walker (4) con valori rispettivamente di $k = 0.00 - 0.03$ nell'accordo stretto e $k = 0.22 - 0.24$ in quello ampliato nel primo studio e di k da -0.27 a 0.36 a seconda del livello spinale nel secondo. Ad esempio, considerando il livello T5-T7 Walker ha riscontrato valori di k pari a 0.12 e accordo interesaminatore al 44% per quanto riguarda l'approccio pragmatico e $k = 0.36$ e accordo al 68% per quanto concerne l'approccio standardizzato.

Anche lo studio di Beynon (24) ha analizzato la mobilità vertebrale secondo un approccio pragmatico e uno standardizzato sia in termini di accordo rigoroso che di accordo regolato alla prevalenza e al rischio di bias PABAK (Prevalence-Adjusted Bias-Adjusted Kappas). L'approccio pragmatico mostra valori di k compresi tra -0.13 e 0.47 per l'accordo rigoroso e PABAK tra -0.14 e 0.71 , l'approccio standardizzato riporta k tra -0.11 e 0.53 per l'accordo rigoroso e PABAK da 0.12 a 0.76 . Nello studio viene anche preso in esame il concetto di vertebra espansa, intesa come la combinazione di tre livelli in uno, basandosi sull'ipotesi che la pressione su un segmento potesse avere un impatto sul livello sopra e sotto e che dal

punto di vista clinico solitamente viene trattata un'area sintomatica piuttosto che un segmento vertebrale specifico. Sulla base di questo l'accordo tra i partecipanti è stato ricalcolato e ha mostrato per l'approccio pragmatico valori di k tra -0.25 e 0.30 per l'accordo rigoroso e PABAK tra 0.09 e 0.89, mentre valori di k tra -0.11 e 0.30 per l'accordo rigoroso e PABAK 0.24 – 0.76 per l'approccio standardizzato.

In questi studi si è evidenziato che gli approcci pragmatici ottengono risultati meno soddisfacenti rispetto a quelli standardizzati.

Cooperstein (19) ha rilevato un buon accordo tra gli esaminatori valutando sia quando essi erano fiduciosi delle proprie misurazioni ($ICC = 0.82$) sia quando entrambi non lo erano ($ICC = 0.39$). Le misurazioni centimetriche ottenute dalla distanza del segmento considerato ipomobile e S1 variavano da una media di 2 cm quando entrambi gli esaminatori erano fiduciosi a una media di 7.10 cm quando nessuno dei due lo era.

L'affidabilità intraoperatore nello studio di Heiderscheit (22) è risultata discreta con $k = 0.17 - 0.26$ secondo l'accordo rigoroso e buona con $k = 0.75 - 0.61$ secondo l'accordo ampliato, con un accordo percentuale pari a 71-85%. Risultati simili sono stati ottenuti da Christensen (23) sia con il disegno hour-to-hour che day-to-day con valori rispettivamente di $k = 0.13 - 0.33$ e di $k = 0.33 - 0.45$ per l'accordo stretto e di $k = 0.60 - 0.68$ e di $k = 0.59 - 0.64$ per l'accordo ampliato.

3.5 Rachide cervicale

Rispetto al rachide lombare e toracico, gli studi che indagano il rachide cervicale includono numerose misurazioni, tra cui le più frequenti sono: ROM attivo, postura, lunghezza e forza muscolare, provocazione dei sintomi e endurance dei muscoli flessori profondi. Nello studio di Piva et al. (12) sono stati riportati anche gli obiettivi secondari, ovvero stabilire il cambiamento minimo rilevabile (MDC) dell'AROM cervicale e determinare l'associazione tra AROM e disabilità.

L'unico studio che si discosta dagli altri è quello di Snodgrass (14) in cui l'obiettivo principale è quello di sviluppare uno strumento affidabile e sicuro per misurare la rigidità segmentale cervicale e documentarla su un gruppo di soggetti asintomatici. Lo strumento è stato utilizzato per misurare simultaneamente l'escursione di un punto del rachide cervicale

applicando una forza meccanica a velocità costante e la resistenza a quella forza, simulando le tecniche di PA con movimenti avanti e indietro a livello del processo spinoso (frequenze di oscillazione da 0,25 a 3Hz, con incrementi di 0,25 Hz).

Per raccogliere i valori di un gruppo di individui, la tensione fornita al motore per tutti i test è stata fissata al 85% del massimo del motore, 10,2 V. L'affidabilità del dispositivo, valutata con misurazioni ripetute su otto diverse combinazioni di schiuma di densità variabile simili all'intervallo di quelle registrate a livello cervicale, è stata valutata con coefficiente di rigidità K.

In questo studio il ruolo del fisioterapista è stato principalmente quello di marcare e pre-condizionare le spinose con 5 PA manuali, come farebbe un clinico per la valutazione, e in seguito lo strumento è stato posto a contatto con la pelle del soggetto e la misurazione è stata presa dopo un'espirazione profonda con frequenza oscillatoria impostata a 1Hz. L'affidabilità tra le misure ripetute è stata calcolata utilizzando il coefficiente di correlazione intra-classe e si è dimostrata buona con ICC (2, 1) = 0,99 (95% CI 0.93 – 1.00).

Questo studio, insieme a quello di Schneider (25) hanno analizzato l'affidabilità intra-operatore. L'affidabilità interoperatore è stata analizzata negli studi di Smedmark (3), Piva (12), Cleland (13), Hanney (15), Schneider (25) e Cooperstein (26).

Gli esaminatori utilizzati in 5 degli studi sono fisioterapisti con diversa specializzazione e esperienza clinica (dai 2 anni di esperienza nello studio di Piva ai 25 nello studio di Smedmark), nello studio di Cooperstein vengono impiegati tre chiropratici con meno di 3 e più di 20 anni di esperienza, mentre nello studio di Hanney si parla di clinici con in media 9 anni di esperienza.

Anche per quanto riguarda il gruppo di soggetti presi in esame non vi è uniformità, dei sette studi solo due prendono in esame soggetti asintomatici (14, 26), mentre i restanti cinque analizzano soggetti sintomatici (3, 12, 13, 15, 25). Il numero dei soggetti studiati varia da un minimo di 22 pazienti negli studi di Hanney e Cleland ad un massimo di 67 nello studio di Snodgrass.

La posizione in cui i soggetti sono stati esaminati varia tra i vari studi e anche all'interno degli stessi studi a seconda del livello preso in esame. Cooperstein e Piva valutano la mobilità segmentale da supino, Cleland valuta da supino i livelli C0-C1 e C1-C2 e da prono il rachide inferiore, Smedmark valuta C1-C2 da seduto, C2-C3 da supino e C7-T1 sul fianco, mentre

Schneider, non considerando C0-C1 e C1-C2, valuta i segmenti da C2 a C7 da prono, così come Snodgrass. Nello studio di Hanney non è specificata la posizione in cui vengono valutati i pazienti.

Solamente 4 studi hanno seguito una procedura di valutazione standardizzata con precedente training (3, 12, 13, 25), mentre negli altri studi non viene riportato nulla a riguardo.

Schneider ha eseguito una valutazione clinica standardizzata ampia analizzando molti aspetti del rachide cervicale oltre al mobilità segmentaria, che lui definisce come MSE manual spinal examination, e ha rilevato valori di affidabilità moderata-eccellente sia intraesaminatore ($k = 0.62 - 0.88$) sia interesaminatore ($k = 0.79 - 0.96$). I livelli che hanno riportato valori maggiori di k sono stati C4-C5 e C5-C6 ($k = 0.82 - 0.88$ intra- e $k = 0.87 - 0.96$ inter-). Questo studio non prende in considerazione C0-C1 e C1-C2, mentre questi livelli sono analizzati nei restanti studi. Cleland (13) ha rilevato valori ponderati di k per la mobilità della colonna vertebrale che andavano da -0.26 a 0.74 (migliori a livello toracico che cervicale con variazione tra lieve e moderato) ed in particolare: C0-C1 $k = -0.26 - 0.46$ e accordo 23-77%, C1-C2 $k = 0.72 - 0.74$ e accordo 86-91%, C2-C7 $k = 0.01 - 0.54$ e accordo 41-77%, T1-T9 $k = 0.13 - 0.82$ e accordo 72-90%.

Smedmark (3) ha dimostrato un accordo totale del 77% e valori di k compresi tra 0,28 e 0,43 considerati solo "equi-moderati". Maggiore accordo a livello di C1-C2 ma k basso (0.28), C2-C3 è risultato il segmento più difficile da valutare ma $k = 0.43$, mentre $k = 0.30$ a livello di C7-T1.

Hanney (15) ha ottenuto valori di k ponderati che vanno da 0.05 a 0.52. Il livello vertebrale C4-5 ha dimostrato il più alto livello di affidabilità ($K = 0.43$), inoltre sono stati considerati separatamente i valori per la regione superiore e inferiore, separate a causa delle loro differenze biomeccaniche intrinseche. L'affidabilità nella colonna vertebrale cervicale superiore ha dimostrato un valore di $k = 0.23$ e nella inferiore $k = 0.30$. L'affidabilità per la valutazione della mobilità dell'articolazione C0-C1 è stata riportata con valori di $k = -0.26 - 0.46$ da Cleland (13), mentre in questo studio k assume valori a partire da 0.15, per la lateroflessione di C2-C3 nel presente studio $k = 0.22 - 0.43$, mentre nello studio di Smedmark (3) $k = 0.43$.

Piva (12) ha riportato un'affidabilità quasi perfetta per la valutazione di C0-C1 con $k = 0.81$ e accordo percentuale del 93%, affidabilità da sufficienti a moderate per quanto riguarda C1-

C2 con $k = 0.30$ e accordo del 64% e per i rachide cervicale inferiore con valori che diminuiscono progredendo da C2 ($k = 0.46$ e accordo 76%) a C7 ($k = -0.07$).

Nello studio di Cooperstein (26) l'accordo globale sul lato dell'ipomobilità è stato del 44%. Rapportando i valori all'intero pool di partecipanti non stratificati ICC [2, 1] = 0,61, con accordo $k = -0.39$ tra il primo e secondo esaminatore, $k = -0.28$ tra il primo e terzo esaminatore e $k = -0.19$ tra il secondo e il terzo esaminatore. La percentuale di partecipanti per cui almeno un esaminatore mancava di fiducia era del 28%, 17% e 11% rispettivamente per gli esaminatori 1, 2 e 3.

4 – DISCUSSIONE

Le tecniche di motion palpation sono parte integrante della maggior parte delle valutazioni utilizzate dai fisioterapisti per analizzare la mobilità del rachide unitamente alla misurazione della mobilità attiva (AROM) e della dolorabilità alla palpazione e al movimento. I movimenti intervertebrali passivi (PIM) vengono valutati in termini di escursione (analisi quantitativa in cui viene stimata la quantità di movimento) e di end-feel (analisi qualitativa in cui i risultati sono interpretati in termini di carattere unisegmentale del movimento) e possono fare da guida nella formulazione del piano di trattamento per le disfunzioni a carico della colonna vertebrale.

In alcuni studi, in particolare in quelli di Cooperstein (19, 20, 26), l'obiettivo era quello di localizzare la vertebra più “fissata”, dove con il termine fissata si intende una vertebra che viene percepita dall'esaminatore esporre più resistenza al movimento dato da una pressione rispetto ad altre vertebre in un intervallo definito. (27)

Nella maggior parte degli studi analizzati, il giudizio della mobilità segmentale si basa sulla mobilità passiva del segmento spinale testato rispetto ai segmenti adiacenti e sull'aspettativa dell'esaminatore. Spesso viene utilizzata una scala a tre punti:

1. Mobilità aumentata: più movimento di quanto normalmente previsto tra il livello testato e i segmenti adiacenti;
2. Mobilità normale: il movimento passivo del livello spinale rientra nei limiti normalmente previsti;
3. Ipomobilità: meno movimento di quanto normalmente previsto tra il livello testato e i segmenti adiacenti.

Nello studio di Schneider (25) la restrizione di movimento è stata classificata come normale, leggera, moderata o marcata, mentre in altri studi la mobilità segmentale è stata compressa per ogni paziente in un rating dicotomico: presenza o assenza di ipomobilità. Ridurre il numero di potenziali rating tende a diminuire l'affidabilità, ma riflette in modo più accurato le valutazioni effettuate nella pratica clinica. Nella pratica clinica se un segmento viene giudicato ipomobile può beneficiare di un qualche tipo di trattamento di mobilizzazione. Viceversa, se in un segmento si riscontra un aumentata mobilità può essere utilizzato un approccio di stabilizzazione. Lo studio di Hicks et al. (18) ha, infatti, riscontrato che

un'aumentata mobilità segmentale lombare e la positività al prone instability test erano associati più frequentemente a quei pazienti che avevano una risposta favorevole agli esercizi di stabilizzazione. Heiderscheit (22) afferma che l'approccio indicato da Hicks risulta coerente con le regole di previsione clinica della manipolazione spinale in pazienti con LBP che includono ipomobilità segmentale spinale come uno dei criteri. L'applicazione appropriata della manipolazione spinale alle aree di "disfunzione spinale" è pensata per migliorare la funzione e il movimento segmentale, con conseguente riduzione del dolore e dei sintomi associati.

Dato il comune utilizzo di queste tecniche negli anni numerosi studi hanno tentato di valutare l'affidabilità della mobilità inter-segmentale della colonna portando a scarsi risultati spesso contraddittori. Nonostante le informazioni provenienti da molti studi sull'affidabilità dimostrino che le tecniche di motion palpation siano inaffidabili, il livello di concordanza tra gli esaminatori suggerisce una leggera affidabilità interesaminatore e moderata a livello intraesaminatore.

Nel dettaglio, l'affidabilità intraoperatore è stata studiata da un solo studio a livello lombare e ha mostrato valori da sufficienti a moderati a seconda del livello che veniva preso in considerazione (17). A livello toracico gli studi che analizzano l'affidabilità intraoperatore sono due (22, 23) e riportano valori di affidabilità da lieve a sufficiente quando viene considerato l'accordo rigoroso e, considerando l'accordo ampliato, i valori salgono a moderati – sostanziali. Christensen ha analizzato l'affidabilità intraoperatore in due modalità diverse ottenendo valori di k accettabili per l'accordo rigoroso sia dell'approccio hour-to-hour che di quello day-to-day e sostanziali per l'accordo ampliato. Date le lievi differenze nelle misurazioni, sembrerebbe che ci sia una variazione biologica nei soggetti nello studio hour-to-hour o che la rilocalizzazione di uno specifico segmento vertebrale è più facilmente eseguita nello stesso invece che in giorni diversi.

Valori decisamente maggiori sono stati riportati nello studio di Schneider (25) per quanto riguarda il rachide cervicale in cui l'affidabilità intraoperatore risulta sostanziale.

I valori di k per l'affidabilità interoperatore lombare risultano sufficienti in 5 studi (11, 16, 17, 18, 20) e sostanziali nello studio di Landel (10). Questi valori salgono a sostanziali quando viene considerato il PABAK e da sufficienti a moderati quando i valori ICC vengono distribuiti in modo omogeneo.

Gli studi che prendono in considerazione la mobilità toracica hanno evidenziato nella maggior parte degli studi bassi valori di affidabilità con frequenti valori negativi (4, 24) o di poco superiori allo zero. Lo studio che ha riportato valori di affidabilità interoperatore maggiori è stato quello di Brisme'e (21) in cui essi andavano da sufficienti a moderati. Quando si parla di accordo ampliato o di PABAK i valori salgono da sufficienti a moderati e, in alcuni casi, anche sostanziali. Cooperstein (19) ha riportato maggiore affidabilità quando tutti e due gli esaminatori erano fiduciosi della misurazione, accordo sostanziale, rispetto a quando nessuno dei due lo era, accordo lieve.

A livello cervicale sono stati studiati segmenti diversi data la loro differente morfologia e funzione. Per quanto riguarda la mobilità del segmento C0-C1 l'affidabilità risulta molto varia con valori che vanno da scarsa (13) a eccellente (12), il segmento C1-C2 ha riportato valori di affidabilità da scarsa a buona. I risultati per il rachide cervicale ottenuti da questa revisione sono in accordo con la revisione di Jonsson et al. del 2017 (28) in cui gli autori hanno esaminato l'affidabilità intra e interoperatore dei test per valutare il movimento segmentale e il dolore in pazienti che presentavano dolore al collo.

Tendenzialmente in tutti i distretti presi in esame l'affidabilità interoperatore risulta consistentemente minore di quella intraoperatore, ad eccezione dello studio di Schneider (25) in cui sono stati ottenuti risultati maggiori per quella interoperatore. Nello studio di Christensen (23), invece, sono stati ottenuti valori di k simili per l'affidabilità intra e interosservatore, ciò fa pensare che la fonte di disaccordo sia legata meno alla soggettività di osservatori diversi rispetto all'incertezza di ogni singolo osservatore.

È ragionevole affermare che la combinazione di diversi test passivi si traduce in un più alto grado di affidabilità per la valutazione di una mobilità segmentale totale rispetto a un test per la valutazione di un particolare movimento segmentale. (16, 25) Inoltre si è osservato che, in accordo con la letteratura precedente, quando viene considerata la provocazione del dolore, intesa come sensazione diversa dalla semplice pressione, sia locale che riferito insieme alla mobilità, i valori di affidabilità risultano più elevati e quando raggruppati sono in grado di fornire un valore decisionale clinico predittivo.

Negli studi analizzati (di cui 6 ad alto rischio di bias, 5 a moderato rischio e 8 a basso rischio) sono numerosi quelli in cui si riscontrano carenze metodologiche diverse, ad esempio in alcuni viene a mancare chiarezza rispetto alla cecità degli esaminatori riguardo alle condizioni cliniche dei soggetti, non sempre le valutazioni vengono eseguite secondo una

precisa randomizzazione e nella gran parte non è specificato il tempo intercorso tra una misurazione e l'altra. Queste carenze abbassano il livello di accuratezza ed esitano in una scarsa affidabilità.

Inoltre, in tutti gli studi, ad eccezione dello studio di Landel et al. in cui veniva utilizzata la risonanza magnetica dinamica per studiare la validità delle tecniche di mobilità segmentale (10), è mancato un Reference Standard.

Possibili spiegazioni per la povera affidabilità generale interesaminatore delle tecniche di motion palpation possono essere ritrovate anche nelle diverse procedure utilizzate (paziente prono/seduto/in decubito laterale, tecniche in PA/lateral glide/flesso-estensione o 3D) e nella scarsa localizzazione del livello spinale che può portare a discrepanze durante le misurazioni. Alcuni autori, per ovviare al problema di una diversa numerazione del livello spinale analizzato, hanno utilizzato dei marker cutanei, applicati preventivamente da un assistente di ricerca, per diminuire il rischio di errore tra le misurazioni dei diversi esaminatori.

Anche l'eterogeneità nel campionamento della popolazione in esame (soggetti sani, sintomatici o mista) può essere in parte causa dei diversi valori di affidabilità riscontrati. È possibile che, qualora vengano analizzati soggetti sintomatici, si possa avere una maggiore variabilità intrinseca all'interno dei soggetti, mentre analizzando soggetti asintomatici potrebbe esserci meno eterogeneità tra le misurazioni non essendo coinvolto l'aspetto soggettivo del dolore. Heiderscheit et al. (22) riportano che, nel caso del rachide toracico, l'uso di soggetti asintomatici non è rilevante, in quanto la valutazione della mobilità toracica è frequentemente eseguita su pazienti privi di dolore alla schiena ma sintomatici nelle regioni adiacenti del corpo (cervicale, cingolo scapolo-omerale).

Secondo lo studio di Johansson et al. (16) le capacità individuali, l'istruzione e l'esperienza dell'esaminatore sono fondamentali per l'esito dei test di mobilità intersegmentale. Nel loro studio, nonostante il livello di istruzione OMT, non sono riusciti a produrre risultati affidabili nelle loro valutazioni di mobilità, mentre altri studi, in cui gli esaminatori avevano lavorato insieme per anni avendo la possibilità di standardizzare la tecnica, hanno ottenuto risultati migliori. (10, 11)

In questa revisione, tuttavia, non è stato possibile confermare quanto riportato nello studio di Johansson, in quanto il livello di esperienza clinica dell'esaminatore non sembra influenzare i risultati.

Si è osservato che anche l'esecuzione di un training prima delle effettive valutazioni con lo scopo di standardizzare la procedura non influenza l'affidabilità dei test, a conferma di ciò i valori di k ottenuti con l'approccio pragmatico non si discostano di molto da quelli ottenuti con l'approccio standardizzato (ad eccezione dello studio di Walker et al.). Inoltre, una situazione in cui sia possibile standardizzare la procedura è poco realistica e difficilmente applicabile alla realtà clinica. I test presi in esame si sono dimostrati utili ma non è possibile definirli affidabili a livello interoperatore in circostanze normali dove non vi è standardizzazione. Al contrario, nello studio di Smedmark (3) si afferma l'importanza di standardizzare le procedure d'indagine siccome il termine "rigidità" può avere varie interpretazioni tra i vari terapisti ed è possibile che la mancanza di omogeneità sia una causa dei bassi valori riscontrati per l'affidabilità interoperatore.

Inoltre, a rendere difficile un confronto dei risultati è anche la notevole differenza dei valori statistici con cui gli studi riportano i dati (k , k pesata, percentuale di accordo e ICC). Quasi la totalità degli studi, ad eccezione di quelli di Cooperstein, utilizzano l'indice di concordanza di Cohen (k) per valutare l'affidabilità. I valori di k per i test di mobilità segmentale hanno mostrato un'affidabilità generalmente da bassa a scarsa, in parte perchè confusi dall'elevato numero di accordi negativi o positivi e dal fatto che il coefficiente k viene influenzato dalla prevalenza: se questa è alta, l'accordo di probabilità è elevato e il kappa è ridotto di conseguenza. (12, 16) Per tenere conto di queste potenziali fonti di errore e migliorare l'interpretazione delle statistiche k in alcuni studi è stato preso in considerazione anche il coefficiente k regolato dalla prevalenza e da bias (PABAK). (4, 11, 24)

Tradizionalmente l'accordo tra due osservatori è definito come "accordo rigoroso" quando vi è una concordanza esatta tra i risultati, ovvero quando i due esaminatori effettuano la stessa constatazione positiva esattamente nello stesso segmento. Tuttavia, nel caso di dolore meccanico del rachide le decisioni cliniche non dipendono dalla palpazione di un singolo segmento, ma di una zona sintomatica. Dal punto di vista clinico e nella decisione sulla terapia più adatta, la numerazione di un certo livello segmentale non è essenziale fintanto che l'area da trattare sia correttamente identificata manualmente. (12) Per questo motivo in molti studi viene indagato anche l'accordo "ampliato" in cui gli esaminatori concordavano su un livello più che su uno specifico segmento. Considerando l'accordo ampliato si sono osservati valori di k superiori rispetto a quelli ottenuti con l'accordo rigoroso.

Per quanto riguarda il valore di ICC, esso diminuisce quando la variabilità della popolazione in oggetto è relativamente bassa, ovvero i soggetti sono relativamente omogenei. Nonostante negli studi di Cooperstein (19, 20, 26) vi sia omogeneità tra i soggetti analizzati con conseguente possibile depressione dell'indice ICC, gli altri tre indici utilizzati hanno suggerito che gli esaminatori in media tendevano a identificare lo stesso livello vertebrale o segmento di movimento come il più fissato.

In questi studi è stato studiato anche il livello di fiducia di ogni esaminatore, quando ogni esaminatore era "molto fiducioso" sul segmento più fissato i livelli identificati erano molto vicini con conseguente accordo "buono". Nello studio riguardante il livello lombare (20) l'affidabilità interesaminatore era più bassa quando entrambi gli esaminatori erano fiduciosi rispetto a quando almeno uno mancava di fiducia. Questi risultati ICC differiscono da quelli ottenuti nello studio riguardante il movimento toracico e quello cervicale in cui la fiducia condivisa dell'esaminatore era associata ad una maggiore affidabilità interesaminatore. Si tratta di risultati non riscontrati in molti altri studi di motion palpation che fa pensare che l'utilizzo di una metodologia di misure continue può essere più pratico per individuare l'accordo tra esaminatori rispetto all'utilizzo di analisi discrete di livello per livello.

Anche nello studio di Snodgrass et al. (14) per il rachide cervicale si sono ottenuti valori diversi rispetto a quelli osservati nella colonna lombare. I principali risultati di questo studio sono stati che la colonna vertebrale cervicale risponde in modo diverso alla forza meccanica rispetto alla colonna vertebrale lombare, con conseguente valori di rigidità inferiori, probabilmente dovuto al fatto che in quest'ultima la respirazione e le contrazioni muscolari possono influenzare le misurazioni.

In conclusione, in accordo con la precedente letteratura (29, 30, 31) questa revisione conferma la scarsa affidabilità della valutazione della mobilità segmentale vertebrale e dimostra la grande importanza di portare avanti lavori futuri con un buon livello metodologico, in cui vengano definiti con più chiarezza le procedure, la popolazione in oggetto e gli indici statistici da utilizzare al fine di ottenere dati più facilmente confrontabili tra loro.

Limiti della revisione

In questa revisione sono presenti alcune limitazioni metodologiche dovute alla restrizione linguistica e alla ridotta ricerca bibliografica su due soli database, benchè poi ampliata mediante l'utilizzo della ricerca manuale, per cui non siano stati trovati alcuni studi pertinenti. Inoltre, la scelta di includere nei criteri di inclusione dello studio solo le tecniche per l'individuazione del segmento ipomobile e non la provocazione del dolore o la localizzazione dei segmenti vertebrali , può aver portato alla perdita di alcuni articoli pertinenti.

5 – CONCLUSIONI

Gli studi di alta qualità per i test di mobilità manuale sono molto rari. La conoscenza di quali strutture vengono interessate durante i test influisce sull'uso di ciascuna tecnica. Tuttavia, nonostante la mancanza di tale conoscenza che crea insicurezza su ciò che viene stimato esattamente (validità), i risultati degli studi sull'affidabilità sono fondamentali per i terapisti manuali, soprattutto se uniti ai risultati ottenuti con la storia clinica del paziente e con il trattamento. I test passivi di movimento intersegmentale vengono anche utilizzati come risorsa aggiuntiva per l'individuazione del livello spinale corretto, tuttavia l'utilizzo di questo metodo per identificare il corretto livello della colonna vertebrale rispetto ad altri metodi più tradizionali non è raccomandato. (32)

Dalla revisione condotta emergono bassi valori sia per quanto riguarda l'affidabilità intraoperatore che per quella interoperatore. Quest'ultima può essere dovuta al fatto che ogni terapeuta ha i propri metodi personali per valutare la mobilità segmentale, pertanto, mancando in molti studi una standardizzazione della procedura, può risultare difficile confrontare le interpretazioni dei diversi esaminatori. Infatti, risulta difficile fare un confronto diretto dei risultati di affidabilità per la mobilità segmentaria e la valutazione dell'ipomobilità tra i vari studi perché essi differiscono molto in termini di tecniche utilizzate, popolazioni analizzate, esaminatori impiegati e indici statistici utilizzati. Tuttavia i risultati suggeriscono che l'affidabilità delle valutazioni della mobilità, indipendentemente dal grado di standardizzazione, sono contrastanti e spesso scarsi o non migliori del caso. Questi risultati mettono in discussione l'uso continuo della palpazione del movimento come parte della valutazione della mobilità vertebrale. A conferma di ciò, nello studio di Landel et al. la valutazione manuale della mobilità segmentale lombare non concorda con il movimento del piano sagittale misurato dalla risonanza magnetica dinamica. Questo risultato mette in discussione la validità della procedura PA come metodo per valutare il movimento intervertebrale della colonna lombare. È possibile che i fisioterapisti basino le loro valutazioni manuali sulla rigidità percepita invece che sul movimento intervertebrale; tuttavia, sono necessarie ulteriori ricerche per testare questa ipotesi.

Altri studi, al contrario, suggeriscono che le valutazioni manuali della mobilità spinale possono essere sufficientemente affidabili per essere un componente utile dell'esame clinico.

Nonostante i dati di accuratezza e affidabilità siano poco incoraggianti, non significa che si debba eliminare le tecniche di motion palpation dalle procedure comunemente utilizzate nella valutazione della mobilità del rachide. Occorre cambiare il ragionamento di fondo per cui queste tecniche vengono utilizzate, inserendole in un approccio più globale per concentrare l'attenzione non sul segmento ipomobile o di dolore ma sui disturbi del paziente. Nella pratica clinica, infatti, l'esame obiettivo comprende sempre più tecniche di valutazione e le conclusioni non sono mai basate solo sui risultati ottenuti da un singolo test. Risulta fondamentale integrare i risultati ottenuti dai test di mobilità segmentale con le informazioni derivanti dall'anamnesi e dal resto della valutazione funzionale, al fine di impostare il trattamento più adatto e coerente alla situazione clinica del paziente.

Risulta, quindi, che tutti i futuri studi di affidabilità abbiano una qualità metodologica superiore e un basso rischio di bias, al fine di ottenere risultati più omogenei e più facilmente alla realtà clinica.

BIBLIOGRAFIA

1. Abbott JH, Flynn TW, Fritz JM, Hing WA, Reid D, Whitman JM. Manual physical assessment of spinal segmental motion: Intent and validity. *Man Ther* [Internet]. 2009;14(1):36–44. Available from: <http://dx.doi.org/10.1016/j.math.2007.09.011>
2. Huijbregts PA. Spinal Motion Palpation: A Review of Reliability Studies. *J Man Manip Ther* [Internet]. 2002;10(1):24–39. Available from: <http://www.tandfonline.com/doi/full/10.1179/106698102792209585>
3. Smedmark V, Wallin M, Arvidsson I. Inter-examiner reliability in assessing passive intervertebral motion of the cervical spine. *Man Ther*. 2000;5(2):97–101.
4. Walker BF, Koppenhaver SL, Stomski NJ, Hebert JJ. Interrater Reliability of Motion Palpation in the Thoracic Spine. *Evidence-Based Complement Altern Med*. 2015;2015:1–6.
5. Pool JJ, Hoving JL, De Vet HC, Van Mameren H, Bouter LM. The interexaminer reproducibility of physical examination of the cervical spine. *J Manipulative Physiol Ther*. 2004;27(2):84–90.
6. Owens EF, DeVocht JW, Wilder DG, Gudavalli MR, Meeker WC. The Reliability of a Posterior-to-Anterior Spinal Stiffness Measuring System in a Population of Patients With Low Back Pain. *J Manipulative Physiol Ther*. 2007;30(2):116–23.
7. Ingram LA, Snodgrass SJ, Rivett DA. Comparison of cervical spine stiffness in individuals with chronic nonspecific neck pain and asymptomatic individuals. *J Orthop Sports Phys Ther*. 2015 Mar;45(3):162-9. doi: 10.2519/jospt.2015.5711. Epub 2015 Jan 27.
8. Zaki R, Bulgiba A, Nordin N, Ismail NA. A systematic review of statistical methods used to test for reliability of medical instruments measuring continuous variables. *Iran J Basic Med Sci*. 2013;16(6):803–7.
9. Lucas N, Macaskill P, Irwig L, Moran R, Rickards L, Turner R, et al. The reliability of a quality appraisal tool for studies of diagnostic reliability (QAREL). *BMC Med Res Methodol* [Internet]. 2013;13(1):1. Available from: BMC Medical Research Methodology.

10. Landel R, Kulig K, Fredericson M, Li B, Powers CM. Intertester Reliability and Validity of Motion Assessments During Lumbar Spine Accessory Motion Testing. *Phys Ther* [Internet]. 2008;88(1):43–9. Available from: <https://academic.oup.com/ptj/ptj/article/2747228/Intertester>
11. Schneider M, Erhard R, Brach J, Tellin W, Imbarlina F, Delitto A. Spinal Palpation for Lumbar Segmental Mobility and Pain Provocation: An Interexaminer Reliability Study. *J Manipulative Physiol Ther* [Internet]. 2008;31(6):465–73. Available from: <http://www.sciencedirect.com/science/article/pii/S0161475408001760>
12. Piva SR, Erhard RE, Childs JD, Browder DA. Inter-tester reliability of passive intervertebral and active movements of the cervical spine. *Man Ther*. 2006 Nov;11(4):321-30. Epub 2006 Jul 11.
13. Cleland JA, Childs JD, Fritz JM, Whitman JM. Interrater Reliability of the History and Physical Examination in Patients With Mechanical Neck Pain. *Arch Phys Med Rehabil*. 2006;87(10):1388–95.
14. Snodgrass SJ, Rivett DA, Robertson VJ. Measuring the posteroanterior stiffness of the cervical spine. *Man Ther*. 2008;13(6):520–8.
15. Hanney WJ, George SZ, Kolber MJ, Young I, Salamh PA, Cleland JA. Inter-rater reliability of select physical examination procedures in patients with neck pain. *Physiother Theory Pract*. 2014;30(5):345–52.
16. Johansson F. Interexaminer reliability of lumbar segmental mobility tests. *Man Ther*. 2006;11(4):331–6.
17. Deore M, May S. The inter-rater and intra-rater reliability of passive physiological accessory movement assessment of lumbar spine in novice manual therapists. *J Bodyw Mov Ther* [Internet]. 2012;16(3):289–93. Available from: <http://dx.doi.org/10.1016/j.jbmt.2011.12.006>
18. Hicks GE, Fritz JM, Delitto A, Mishock J. Interrater reliability of clinical examination measures for identification of lumbar segmental instability. *Arch Phys Med Rehabil*. 2003;84(12):1858–64.
19. Cooperstein R, Haneline M, Young M. Interexaminer reliability of thoracic motion palpation using confidence ratings and continuous analysis. *J Chiropr Med* [Internet].

- 2010;9(3):99–106. Available from: <http://www.sciencedirect.com/science/article/pii/S1556370710000702>
20. Cooperstein R, Young M. The reliability of lumbar motion palpation using continuous analysis and confidence ratings: choosing a relevant index of agreement. *J Can Chiropr Assoc* [Internet]. 2016;60(2):146–57. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27385834><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4915474>
 21. Brismée JM, Gipson D, Ivie D, Lopez A, Moore M, Matthijs O, et al. Interrater Reliability of a Passive Physiological Intervertebral Motion Test in the Mid-Thoracic Spine. *J Manipulative Physiol Ther*. 2006;29(5):368–73.
 22. Heiderscheit B, Boissonnault W, et al. Reliability of Joint Mobility and Pain Assessment of the Thoracic Spine and Rib Cage in Asymptomatic Individuals. *J Man Manip Ther*. 2008;16(4):210–6.
 23. Christensen HW, Vach W, Vach K, Manniche C, Haghfelt T, Hartvigsen L, et al. Palpation of the upper thoracic spine: An observer reliability study. *J Manipulative Physiol Ther*. 2002;25(5):285–92.
 24. Beynon AM, Hebert JJ, Walker BF. The interrater reliability of static palpation of the thoracic spine for eliciting tenderness and stiffness to test for a manipulable lesion. *Chiropr Man Ther*. 2018;26(1):1–10.
 25. Schneider GM, Jull G, Thomas K, Smith A, Emery C, Faris P, et al. Intrarater and interrater reliability of select clinical tests in patients referred for diagnostic facet joint blocks in the cervical spine. *Arch Phys Med Rehabil* [Internet]. 2013;94(8):1628–34. Available from: <http://dx.doi.org/10.1016/j.apmr.2013.02.015>
 26. Cooperstein R, Young M, Haneline M. Interexaminer reliability of cervical motion palpation using continuous measures and rater confidence levels. *J Can Chiropr Assoc* [Internet]. 2013;57(2):156–64. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23754861><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3661183>
 27. Cooperstein R, Young M. The reliability of spinal motion palpation determination of the location of the stiffest spinal site is influenced by confidence ratings: A secondary

- analysis of three studies. *Chiropr Man Ther* [Internet]. 2016;24(1):1–13. Available from: <http://dx.doi.org/10.1186/s12998-016-0131-x>
28. Jonsson A, Rasmussen-Barr E. Intra- and inter-rater reliability of movement and palpation tests in patients with neck pain: A systematic review. *Physiother Theory Pract* [Internet]. 2018;34(3):165–80. Available from: <https://doi.org/10.1080/09593985.2017.1390806>
 29. Stochkendahl MJ, Christensen HW, Hartvigsen J, Vach W, Haas M, Hestbaek L, et al. Manual examination of the spine: a systematic critical literature review of reproducibility. *J Manipulative Physiol Ther*. 2006;29(6):475–485.e10.
 30. Tuttle N, Hazle C. Spinal PA movements behave ‘as if’ there are limitations of local segmental mobility and are large enough to be perceivable by manual palpation: A synthesis of the literature. *Musculoskelet Sci Pract* [Internet]. 2018;36:25–31. Available from: <http://www.sciencedirect.com/science/article/pii/S2468781218301280>
 31. Van Trijffel E, Anderegg Q, Bossuyt PMM, Lucas C. Inter-examiner reliability of passive assessment of intervertebral motion in the cervical and lumbar spine: A systematic review. *Man Ther*. 2005;10(4):256–69.
 32. Chakraverty RC, Pynsent PB, Westwood A, Chakraverty JK. Identification of the correct lumbar level using passive intersegmental motion testing. *Anaesthesia*. 2007;62(11):1121–5.