



UNIVERSITÀ DEGLI STUDI  
DI GENOVA



## **Università degli Studi di Genova**

Facoltà di Medicina e Chirurgia

### **Master in Riabilitazione dei Disordini Muscoloscheletrici**

A.A. 2011-2012

Campus Universitario di Savona

# **Accuratezza diagnostica nell'individuazione di problematiche di intrappolamento della radice nervosa nel mal di schiena acuto.**

Candidato:

Vito Curri

Relatore

Luca Viganò

# INDICE

## ABSTRACT

### 1. INTRODUZIONE

- 1.1. Low back Pain: principali caratteristiche
- 1.2. Radicolopatia: elementi di patofisiologia
- 1.3. Sindrome radicolare: esame clinico

### 2. TEST DIAGNOSTICI

- 2.1. Qualità psicometriche di un test: Accuratezza Diagnostica, Sensibilità, Specificità, Valore predittivo positivo e negativo, F1-Score
- 2.2. Risonanza Magnetica: Gold Standard
- 2.3. Test di valutazione della sindrome radicolare nel LBP acuto

### 3. OBIETTIVI DELLO STUDIO

### 4. MATERIALI E METODI

- 4.1. Fonti di informazione
- 4.2. Stringhe di ricerca
- 4.3. Criteri di inclusione ed esclusione
- 4.4. Scala Quadas e F1 Score

### 5. RISULTATI

- 5.1. Diagramma di flusso: selezione degli articoli
- 5.2. Analisi descrittiva degli studi oggetto della revisione
- 5.3. Valutazione qualità metodologica degli studi: SCALA QUADAS
- 5.4. Valutazione qualità psicometriche dei test: TABELLA F1-SCORE

### 6. DISCUSSIONE

### 7. CONCLUSIONI

### 8. KEY POINTS

### 9. BIBLIOGRAFIA

## ABSTRACT

**Obiettivi:** questo studio si propone, in una prima fase, l'obiettivo di analizzare, attraverso la *Scala Quadas*, la qualità metodologica degli studi presenti in letteratura sull'accuratezza diagnostica dei principali test per la valutazione della sindrome radicolare nel LBP acuto. Successivamente, i test clinici relativi agli studi con valutazione Quadas più alta verranno valutati attraverso l'applicazione dell'F1 Score, al fine di individuare il test che presenta il miglior valore di accuratezza diagnostica.

**Materiali e Metodi:** La ricerca è stata condotta consultando la banca dati elettronica Medline. I termini chiave utilizzati sono stati i seguenti: (((("straight leg raise test" OR "straight leg raising test" OR "slr test" OR "crossed straight leg raising test" OR "lasègue" OR "lasègue's" OR "lasègue test" OR "lasègue's test" OR "lasègue sign" OR "lasègue's sign" OR "slump" OR "slump test" OR "kemp" OR "kemp's" OR "fingertip to floor test" OR "finger to floor distance" OR "bell" OR "bell's" OR "bell test" OR "X sign" OR "femoral stretch test" OR "femoral nerve stretch test" OR "wassermann" OR "wassermann's" OR "wassermann test" OR "wassermann's test" OR "bragard" OR "bragard's" OR "bragard test" OR "bragard's sign" OR "neri" OR "neri's" OR "knee flexion test" OR "valsalva maneuver" OR "valsalva's maneuver" OR "hyperextension test" OR "bowstring test" OR "bowstring sign"))) AND ("nerve root compression" OR "radiculopathy" OR "sciatica" OR "nerve impingement" OR "low back pain")) AND ("sensitivity" OR "specificity" OR "accuracy" OR "validity" OR "diagnosis" OR "differential" OR "reference standards"). Dalla ricerca sono risultati 152 articoli.

**Criteri di inclusione ed esclusione:** Le tipologie di studio oggetto della presente revisione sono state le seguenti: RCT, studi di coorte, studi caso-controllo, case report. Sono stati presi in considerazione studi in lingua inglese o italiana e con abstract consultabile, che avessero come oggetto la valutazione di pazienti con segni e sintomi di sofferenza radicolare, diagnosticata attraverso Risonanza Magnetica, Tomografia Assiale Computerizzata, Mielografia o tramite riscontro diretto chirurgico.

I criteri di esclusione sono stati i seguenti:

- a. Articoli non in lingua inglese o italiana;
- b. Abstract non disponibile;
- c. Disegno dello studio non appropriato;
- d. Articoli di cui non è stato possibile ricavare il full text;
- e. Articoli non rilevanti ai fini dell'indagine.

**Risultati:** Gli studi validi alla fine della revisione, dopo lettura completa, sono risultati 15. La valutazione con la **Scala Quidas** ha messo in rilievo alcune criticità, relativamente alla qualità metodologica degli studi. Innanzitutto, la scelta del campione condiziona pesantemente i valori di accuratezza diagnostica dei test, ma li rende poco generalizzabili. Il tempo trascorso tra l'esecuzione del test clinico e l'analisi strumentale (golden standard) è un altro aspetto che può inficiare l'attendibilità dei risultati. La totale assenza della descrizione dei metodi di esecuzione ed interpretazione dei test pregiudica la loro ripetibilità in contesti diversi da quello in esame.

Alla valutazione con l'F1 Score, lo **Slump test** è risultato il test dotato della migliore accuratezza diagnostica per le problematiche di sofferenza della radice nervosa in generale e di compressione radicolare in particolare, con un punteggio pari a 0,84 contro il punteggio di 0,64 dello straight leg raise (SLR) test. Quest'ultimo, per quanto rimanga uno degli strumenti più importanti e maggiormente utilizzati nella valutazione delle radicolopatie, è soprattutto un indicatore di tensione e conseguente irritazione della radice nervosa, piuttosto che di compressione radicolare.

**Conclusioni:** La maggior parte degli studi tratta di ernia/protrusione discale e non di sofferenza radicolare. Sarebbe pertanto più corretto inquadrare le future ricerche in un quadro più ampio di sindrome radicolare, a prescindere dai singoli fattori eziologici. Inoltre, non esistono attualmente in letteratura o sono molto pochi gli studi che attestino l'efficacia di diversi test clinici, di utilizzo comune nella pratica clinica. Ciò limita in modo consistente lo spettro di strumenti clinici basati su prove di efficacia a disposizione del fisioterapista, impoverendone il "potenziale diagnostico" e l'individuazione del percorso riabilitativo più efficace per il paziente.

L'anamnesi è il punto di partenza fondamentale per la valutazione e la diagnosi di sindrome radicolare. L'utilizzo in maniera isolata di un test non sempre garantisce dei risultati attendibili. Di conseguenza, la combinazione di più test clinici integrata con un'approfondita fase anamnestica rappresenta senz'altro la migliore "*strategia diagnostica*" per identificare una sindrome radicolare, nell'ambito di un processo di ragionamento clinico più globale.

# 1. INTRODUZIONE

## 1.1. Low back pain: principali caratteristiche

Il low back pain (di seguito LBP) è definito come dolore e/o limitazione funzionale compreso tra il margine inferiore dell'arcata costale e le pieghe glutee inferiori con eventuale irradiazione posteriore alla coscia ma non oltre il ginocchio che può causare l'impossibilità di svolgere la normale attività quotidiana, con possibile assenza dal lavoro<sup>1</sup>.

Tale disturbo osteoarticolare affligge quasi l'80% della popolazione, si presenta in egual misura nei due sessi ed insorge più spesso fra i 30 ed i 50 anni di età<sup>1</sup>.

Uno studio effettuato da Rabin et al. nel 2007 sostiene che 2/3 degli adulti soffriranno di lombalgia (LBP) in qualche momento della loro vita, in particolar modo nel quarto e quinto decennio della loro vita<sup>2</sup>.

Majlesi in uno studio del 2008 mette in evidenza una maggiore frequenza nei soggetti con età inferiore ai 45 anni o che praticano alcuni sport, come ad esempio ginnastica, wrestling, calcio, tennis, golf e danza. La prevalenza di tale patologia si attesta tra il 75% e l'85% mentre l'incidenza è del 5%<sup>3</sup>.

Due studi indicano il LBP come la principale causa di disabilità nei Paesi industrializzati. Tuttavia, nonostante molte persone abbiano almeno un episodio di LBP nella loro vita, in circa l'85% dei pazienti non viene identificata alcuna patologia specifica<sup>4,5</sup>.

Pertanto, in relazione al fattore eziologico, il LBP può distinguersi in specifico (15% dei casi) o aspecifico (85% dei casi). Secondo Majlesi, quest'ultima tipologia presenta alcune ambiguità terminologiche: si intendono tutte quelle situazioni in cui non è facilmente identificabile la fonte del dolore oppure nelle quali non è presente una patologia ben definita? Il più delle volte la diagnosi di LBP aspecifico è fatta "per esclusione", ossia dopo aver escluso tutta una serie di patologie "specifiche", quali ad esempio fratture, ernie, tumori, spondilite anchilosante, artrite reumatoide, stenosi, infezioni, sindrome della cauda equina, aneurisma dell'aorta, ecc. Altro aspetto rilevante nella diagnosi di LBP è la relazione tra la presenza di impairments dal punto di vista "anatomico", come ad esempio bulging o estroflessioni del disco intervertebrale, messe in rilievo attraverso tecniche di imaging e l'effettivo quadro clinico e sintomatologico del paziente. Bisogna porre certamente molta attenzione all'interpretazione della diagnostica strumentale (RM, TC, ecc.) ed alla relativa correlazione con il quadro clinico del paziente, in quanto una grande

proporzione di “anomalie discali”, che includono processi degenerativi e protrusioni, sono state rilevate in soggetti asintomatici<sup>3</sup>.

In relazione al fattore temporale, possiamo distinguere il LBP in acuto, ossia inferiore alle 4 settimane, sub-acuto con durata tra 4 settimane e 3 mesi e cronico, con durata oltre i 3 mesi. Il LBP ricorrente invece, è rappresentato da episodi acuti che durano meno di 12 settimane e si ripresentano dopo un periodo di benessere<sup>1</sup>. Il dolore alla gamba è presente nel 57% dei pazienti con LBP. Questo dolore può derivare dalla compressione di strutture neurali oppure essere un dolore riferito che deriva dalla stimolazione dannosa del disco intervertebrale, delle faccette articolari o dei muscoli a livello lombare<sup>6</sup>.

Il presente lavoro affronterà in particolare la valutazione delle problematiche di intrappolamento della radice nervosa nel LBP acuto.

## 1.2. Radicolopatia: elementi di patofisiologia

Con il termine "radicolopatia" ci si riferisce genericamente ad una sofferenza di una radice nervosa che determina dolore ed eventualmente disturbi motori e sensitivi nei rispettivi territori d'innervazione. I segni ed i sintomi principali di una radicolopatia sono: dolore, parestesie, diminuzione o assenza dei riflessi osteotendinei, deficit di forza e sensibilità. Tali sintomi sono distribuiti in modo univoco all'interno del territorio innervato dalla radice nervosa interessata<sup>2</sup>.

Più specificatamente, a livello lombare, possiamo distinguere la sciatalgia dalla cruralgia. La prima ha una prevalenza stimata intorno al 5%, può essere associata a mal di schiena (lombosciatalgia), è caratterizzata da dolore irradiato lungo il decorso del nervo sciatico, dal gluteo alla parte posteriore della coscia e postero-laterale della gamba, fino alla caviglia. E' importante fare distinzione tra la sciatalgia "vera e propria" e la sciatica, caratterizzata da dolore riferito non di origine radicolare.

La cruralgia è meno comune e causa dolore lungo la faccia anteriore o antero-interna della coscia, ossia lungo il decorso del nervo crurale.

L'eziologia del dolore radicolare è ascrivibile a diversi fattori, quali ad esempio: bulging o ernia discale, stenosi spinale, osteofitosi, processi degenerativi della faccetta articolare, cisti sinoviali, tumori, infezioni ed anomalie congenite<sup>7</sup>.

Nei pazienti che riferiscono i sintomi che si irradiano in gamba (sciatalgia), i medici valutano le possibili cause di radicolopatia (compressione della radice nervosa) attraverso la storia clinica, l'esame fisico e la diagnostica per imaging. Una delle cause di radicolopatia può essere un'ernia del disco, che esercita una pressione sulla radice nervosa. L'ernia del disco si verifica spesso senza sintomi, come rivelato da studi di risonanza magnetica in molte persone asintomatiche. Essi diventano sintomatici quando il materiale discale entra in contatto diretto con la radice nervosa, provocando un'irritazione meccanica e chimica del nervo a quel livello. Il 35% dei soggetti con ernia sviluppano una sindrome radicolare<sup>5</sup>.

I livelli più frequentemente coinvolti dalla radicolopatia sono:

- tra i 25 e 55 anni: L4 –L5 e L5 –S1 ( 95 % dei casi);
- dopo i 55 anni: livelli superiori.

I fattori di rischio per l'irritazione della radice nervosa sono di tre tipi: personali, lavorativi e comportamentali:

**1. PERSONALI:**

- ✓ Età compresa fra 45 e 64 anni (picco);
- ✓ Altezza (> di 185 cm);
- ✓ Stress;

**2. LAVORATIVI:**

- ✓ Frequenti sollevamenti associati a flessioni e torsioni del busto<sup>8</sup>;
- ✓ La prolungata posizione seduta in flessione (più di mezza giornata lavorativa) o rotazione, associata a vibrazioni trasmesse al corpo, è un forte fattore di rischio per lo sviluppo di LBP e/o sciatica; un esempio è costituito dai piloti di elicottero<sup>9</sup>;

**3. COMPORTAMENTALI**

- ✓ Obesità (l'intensità dei sintomi non aumenta proporzionalmente al peso, ma accanto al fattore meccanico è presente un aumento di markers infiammatori, come la Proteina C-reattiva);
- ✓ Fumo da più di venti anni (effetto citotossico della nicotina ed aumento della Proteina C-reattiva);
- ✓ Nelle persone con aumento nel plasma di Proteina C-reattiva ed alti livelli di attività fisica<sup>10</sup>.

### 1.3. Sindrome radicolare: esame clinico

La valutazione funzionale di una sindrome radicolare si avvale della combinazione di più elementi: anamnesi clinica del paziente, esame obiettivo e diagnostica strumentale. Il Golden Standard per la valutazione di una sindrome radicolare è la risonanza magnetica (RMN); in alternativa viene utilizzata la TAC. Nella fase anamnestica e dell'esame obiettivo è innanzitutto necessario escludere le "Red Flags", ossia rilevare segni e sintomi che possono collegare un disordine muscoloscheletrico ad una patologia grave. A livello lombare è necessario escludere, ad esempio, fratture vertebrali, tumori, sindromi della cauda equina, infezioni, aneurismi dell'aorta, lombalgie di origine infiammatoria o viscerale.

L'anamnesi procede con un'approfondita analisi della sintomatologia del paziente (ad es. tipologia, localizzazione, cronologia del dolore), dei movimenti provocativi, delle menomazioni, limitazione di attività e restrizione nella partecipazione in ottica bio-psico-sociale, seguendo il modello ICF. Successivamente vanno individuati i "segni tipici" della sindrome radicolare, in particolare:

- Netta demarcazione della localizzazione (dermatomero)
- Segni neurologici:
  - Sensibilità
  - Riflessi
  - Forza
  - Positività ai test neurodinamici
- Deviazione antalgica (shift rilevabile anche a paziente vestito)
- Perdita di movimenti in flessione e dolore in estensione<sup>11</sup>.

#### Localizzazione Dermatomerica

In presenza di sindrome radicolare, l'irradiazione del dolore segue un decorso di tipo dermatomerico.



## Segni neurologici:

- 1) **Sensibilità:** esistono differenti tipologie di sensibilità che vengono valutate con strumenti differenti. La sensibilità protopatica o “grossolana” si misura con l’algometro, la sensibilità epicritica o “fine” viene misurata con il test del monofilamento e quella profonda detta anche “pallestesica” o “vibratoria” si avvale del diapason per la sua misurazione. Per la valutazione delle sensibilità superficiale (protopatica ed epicritica) ci sono delle regioni cutanee specifiche, definite dall’autore *Nitta Hiroyuki*, che corrispondono alle radici nervose di L4, L5 e S1; in particolare per L4 si stimola la parte mediale e prossimale della gamba; per L5 si stimola la parte mediale del I metatarso e per S1 si stimola il bordo esterno del V metatarso. Anche per la sensibilità profonda esistono specifiche regioni cutanee sulle quali applicare il diapason, nello specifico la rotula per L4, il malleolo mediale per L5 ed il malleolo laterale o la testa del perone per S1<sup>8</sup>.
- 2) **Riflessi Osteo-tendinei:** una compressione radicolare può provocare altresì la riduzione fino alla totale assenza dei riflessi osteo-tendinei. Tuttavia, l’elicitazione del riflesso può essere condizionata da altri fattori, quali ad esempio lo stato di non rilassamento del paziente piuttosto che patologie concomitanti. È perciò utile eseguire un confronto tra i riflessi di un arto e del controlaterale al fine di evidenziarne eventuali differenze. L’esame clinico dei riflessi osteo-tendinei a livello delle radici lombari prevede la valutazione del riflesso patellare per le radici nervose di L3-L4 e di quello achilleo per le radici nervose di L5-S1. Il primo viene elicitato attraverso la percussione del tendine rotuleo con un martelletto neurologico, il secondo attraverso la percussione del tendine d’Achille<sup>12</sup>.

**Tabella 1:** Riflessi osteo-tendinei

<b>RADICE NERVOSA</b>	<b>TIPOLOGIA DI RIFLESSO</b>	<b>MANOVRA PROVOCATIVA</b>
L3-L4	PATELLARE	Percussione del tendine rotuleo
L5-S1	ACHILLEO	Percussione del tendine d’Achille

3) **Forza Muscolare:** la forza muscolare è l'ultima ad essere compromessa a causa di una sindrome radicolare e denota una sofferenza importante della radice nervosa<sup>12</sup>.

**Tabella 2:** muscoli chiave

<b>RADICE PREVALENTE</b>	<b>MUSCOLO INNERVATO</b>	<b>MOVIMENTO EFFETTUATO</b>
L1	ILEO-PSOAS	FLESSIONE ANCA
L2	ADDUTTORI	ADDUZIONE ANCA
L3	QUADRICIPITE	ESTENSIONE GINOCCHIO
L4	TIBIALE ANTERIORE	DORSIFLESSIONE CAVIGLIA
L5	ESTENSORE LUNGO DELL'ALLUCE	EST. I DITO DEL PIEDE EVERSIONE PIEDE
S1	TRICIPITE SURALE	FLESSIONE PLANT. CAVIGLIA
S2	GRANDE GLUTEO	ESTENSIONE ANCA

**Positività ai test neurodinamici:** questi test valutano la risposta del nervo allo stiramento passivo. Qualora la radice nervosa sia irritata da compressione a livello del forame intervertebrale, lo stiramento provocherà il sintomo doloroso irradiato all'arto inferiore, riconosciuto come familiare dal paziente. Non sempre però la sintomatologia evocata dai test di neurotensione scaturisce in maniera univoca dalla messa in tensione delle strutture neurali. Difatti, oltre ad esse, vengono messe in tensione anche le strutture muscoloscheletriche adiacenti al nervo stesso e, pertanto, i sintomi del paziente potrebbero derivare dalla provocazione di queste ultime. Per discriminare la sollecitazione delle strutture neurali da quelle muscoloscheletriche esistono specifiche fasi del test che hanno finalità di differenziazione strutturale. I principali test neurodinamici sono: test di straight leg raise (SLR) e test di straight leg raise crociato (XSLR o CSLR) per la valutazione del nervo sciatico e delle radici nervose da L4 a S3; test di prone knee bending (PKB) per la valutazione degli impairments a livello delle radici nervose di L2, L3 e L4. Nella popolazione destinata alla chirurgia l'SLR ha dimostrato di avere un'alta sensibilità e specificità variabile, mentre l'XSLR ha mostrato un'alta specificità con bassa sensibilità. L'utilizzo di batterie di test, in aggiunta all'anamnesi e all'esame fisico, può migliorare la performance dei test diagnostici<sup>5</sup>.

### **Deviazione antalgica**

Circa la metà dei pazienti con sindrome radicolare presenta un atteggiamento antalgico in shift laterale. L'ipotesi è che si tratti di una contrattura della muscolatura paravertebrale lombare, allo scopo di decomprimere le strutture nervose a contatto con l'ernia. La convessità della curva

dipende dalla posizione, mediale o laterale, dell'ernia rispetto alla radice nervosa compressa. Tale postura antalgica è spesso riscontrabile soprattutto nella fase acuta della patologia<sup>12</sup>.

### **Perdita di movimenti in flessione e dolore in estensione**

In uno studio del 1999 Stankovic ha esaminato 105 pazienti affetti di ernia del disco ed ha rilevato una significativa riduzione del ROM in flessione del tronco ed un aumento del dolore irradiato all'arto inferiore nei movimenti di estensione, partendo dalla stazione eretta. Inoltre, in 72 su 105 pazienti è stato riscontrato accordo tra la diagnosi effettuata tramite imaging e l'esame clinico basato sull'estensione lombare e lo Slump test. Pertanto, la positività a questi ultimi test potrebbe essere considerata utile per la diagnosi di ernia del disco lombare, con un buon livello di affidabilità.

Per riuscire a distinguere tra pazienti con dolore lombosacrale con o senza radicolopatia a causa dell'ernia del disco, i diversi aspetti dell'esame fisico vanno considerati in maniera integrata e non isolata<sup>11</sup>.

### **Percorsi diagnostici terapeutici**

I percorsi diagnostici terapeutici (PDT) sono delle linee guida che ci aiutano nell'individuazione del livello coinvolto dalla radicolopatia, in presenza di determinati deficit sensitivi, motori e dei riflessi osteo-tendinei.

Nella **RADICOLOPATIA DI L3** il test di Wassermann risulterà positivo, i disturbi sensitivi riguarderanno la regione inguinale e la faccia mediale della coscia e di parte della gamba, il riflesso patellare sarà ridotto o addirittura assente e, in presenza di compressione radicolare importante, assisteremo alla riduzione della forza in estensione del ginocchio.

Nella **RADICOLOPATIA DI L4** il test di Wassermann e quello di Lasègue risulteranno entrambi positivi, la sensibilità della faccia anteriore della coscia, della porzione mediale della gamba e del bordo interno del piede sarà alterata, il riflesso patellare si presenterà ridotto o del tutto assente e, dal punto di vista muscolare, avremo una riduzione della forza in dorsiflessione del piede ed estensione del ginocchio.

Nella **RADICOLOPATIA DI L5** il test di Lasègue risulterà positivo, la sensibilità della faccia laterale della coscia, della gamba, del dorso del piede e dell'alluce sarà alterata, il riflesso achilleo sarà presente, la forza in flessione dorsale dell'alluce e delle dita risulterà ridotta.

Nella **RADICOLOPATIA DI S1** il test di Lasègue risulterà positivo, la sensibilità della faccia posteriore della coscia, della gamba e della regione plantare laterale del piede sarà alterata, il riflesso achilleo si presenterà ridotto o completamente assente, la forza in flessione plantare del piede e dell'alluce sarà ridotta.

In presenza di positività degli elementi descritti nei percorsi diagnostico terapeutici ci sarà un'alta probabilità di trovarsi in presenza di lombosciatalgia o cruralgia; viceversa, in caso di negatività, non potremo comunque escludere del tutto la radicolopatia, ma sarà più probabile che si tratti di una lombalgia aspecifica (SI: PDT lombosciatalgia/cruralgia; NO:FORZA B > PDT lombalgia aspecifica)<sup>13</sup>.

**Tabella 3:** prospetto riassuntivo PDT

<b>RADICE NERVOSA</b>	<b>L3</b>	<b>L4</b>	<b>L5</b>	<b>S1</b>
<b>TEST NEUROD.</b>	Wasserman +	Wasser./SLR +	SLR +	SLR +
<b>DEFICIT SENSIBILITA'</b>	Inguine, coscia e parte gamba mediale	Coscia ant., gamba mediale e bordo interno piede	Coscia e gambe laterale, Piede ed alluce dorsale	Coscia e gamba posteriore, faccia plant. later. piede
<b>R.O.T.</b>	Patellare assente	Patellare assente	Achilleo presente	Achilleo assente
<b>FORZA MUSCOL.</b>	Ext ginocchio	Ext gin. + F.dors. piede	Flex dorsale alluce e dita	Flex plantare piede e dita

In uno studio del 2002 Vroomen individua alcuni elementi ricorrenti nei pazienti affetti da compressione della radice nervosa: età compresa tra 51 e 81 anni (OR 2.2), sesso maschile (OR 1.8), tipo di lavoro: stazione eretta prolungata, camminare, sollevare pesi (OR 0.5), dolore prevalente nella gamba rispetto alla schiena (OR 5.5), tipica distribuzione dermatomerica (OR 3.8), aumento del dolore tossendo, starnutando e compiendo sforzi (OR 2.1), sensazione di freddo con distribuzione dermatomerica nella gamba (OR 1.8), presenza di paresi (OR 5.2), finger to floor distance test > o = di 25 cm (OR 2.4), assenza di riflessi tendinei all'anca ed al ginocchio (OR 2.4), positività del SLR test (OR 2.3)<sup>14</sup>.

In sintesi, il protocollo di esame clinico prevede:

- Ispezione postura (valutare shift);

- Test preliminare: è costituito dal Kemp's test e dallo Slump test; se l'esame soggettivo è fortemente indicativo della presenza di una sindrome radicolare, l'esame preliminare è inutile;
- Esame neurologico: se il test preliminare è positivo, l'esame neurologico va fatto sempre per valutare la radicolopatia, viceversa se è negativo bisogna comunque considerarne l'utilizzo.

## 2. TEST DIAGNOSTICI

### 2.1. Qualità psicometriche di un test: Accuratezza Diagnostica, Sensibilità, Specificità, Valore predittivo positivo e negativo, F1-Score

I test diagnostici rientrano nel più complesso processo di valutazione funzionale del paziente che si avvale dell'anamnesi clinica e dell'esame obiettivo, nell'ambito del quale tali test vengono utilizzati in maniera isolata o in batteria. Nella pratica quotidiana ci si pone sempre la domanda se la misurazione che otteniamo può essere considerata affidabile e precisa ed è in grado di modificare la probabilità che un paziente abbia o meno una particolare patologia. I concetti di affidabilità, precisione, accuratezza, sensibilità, specificità, valore predittivo, rapporto di verosimiglianza rappresentano alcune caratteristiche tecniche intrinseche dei test diagnostici, che ci possono aiutare a capire se lo strumento di misurazione utilizzato può supportarci in maniera efficace nella nostra pratica clinica.

#### **AFFIDABILITA'**

Per attendibilità di un test si intende la coerenza fra i punteggi rilevati con esso in due momenti successivi; i punteggi ottenuti devono essere il più possibile indipendenti dalle variazioni accidentali nelle condizioni dei soggetti testati ed in quelle di somministrazione e da fattori che possono costituire fonti di errore nella misurazione.

La validità, invece, rappresenta il grado o la precisione con cui un test rileva una variabile per la misurazione della quale il test è stato costruito.

Secondo Pedrabissi, un test è tanto più affidabile quanto più fornisce misure accurate, cioè immuni da errori casuali. Quante più variazioni nei punteggi si ottengono, a causa di errori casuali fra diverse somministrazioni di uno stesso test allo stesso gruppo di soggetti, tanto meno questi strumenti sono attendibili.

Pertanto, una misurazione è affidabile se i risultati rimangono costanti nel tempo. In questo caso si avrà un'alta affidabilità, mentre uno strumento di misura non è affidabile se le misure ripetute forniscono risultati differenti, nonostante non vi siano cambiamenti evidenti nei soggetti esaminati<sup>15</sup>.

Secondo Cleland, un test è ritenuto affidabile se fornisce informazioni precise, esatte e riproducibili. Quest'ultima proprietà si ottiene attraverso un'attenta e dettagliata descrizione delle

sue procedure operative, in modo tale da poter garantire costanza metodologica. La precisione non deve riguardare soltanto la descrizione, bensì anche le modalità di esecuzione dei test da parte degli esaminatori. Va inoltre *“concordata”* la definizione dei criteri di positività e negatività di un test, che dev’essere coerente tra i vari esaminatori, in modo tale da non inficiarne l’attendibilità.

Possiamo distinguere due tipologie di affidabilità: intra-esaminatore ed inter-esaminatore. La prima è la capacità di ottenere la stessa misura in applicazioni ripetute del test da parte dello stesso esaminatore. La seconda esprime il grado di concordanza tra due o più esaminatori nell’ottenere risultati identici in applicazioni ripetute dello stesso test<sup>16</sup>.

### **ACCURATEZZA DIAGNOSTICA**

Per accuratezza di un esame diagnostico si intende la sua capacità di fornire valori corrispondenti a quelli veri. Pertanto, l’accuratezza diagnostica scaturisce dal confronto tra il “test clinico” ed un “Golden Standard”, che rappresenta il metodo più accurato e più vicino alla realtà per identificare una malattia. Maggiore è l’accordo tra i risultati ottenuti con il test clinico e quelli ottenuti con il golden standard, migliore sarà l’accuratezza diagnostica del test in esame. L’accuratezza diagnostica è legata strettamente ai concetti di: **“Valore Predittivo Positivo” (PPV)** e **“Valore Predittivo Negativo” (NPV)**, **“Sensibilità”**, **“Specificità”**, **“rapporto di verosimiglianza”**, **“Curva di ROC”** e **“F1 SCORE”**<sup>17</sup>.

Per comparare i risultati ottenuti applicando il test diagnostico con quelli ottenuti con il Golden Standard si utilizza una *tabella di contingenza 2x2*, suddivisa in 4 sezioni, in modo tale da determinare la capacità di un test di identificare correttamente i veri positivi, i veri negativi, i falsi positivi (in cui i test diagnostici sono positivi, mentre il Golden Standard è negativo) ed i falsi negativi (in cui i test diagnostici sono negativi, mentre il Golden Standard è positivo).

**Tabella 4: Tabella di contingenza 2x2 per comparazione risultati Test-Golden STD**

<b>TEST DIAGNOSTICO</b>	<b>GOLDEN STANDARD POSITIVO</b>	<b>GOLDEN STANDARD NEGATIVO</b>
Test Clinico Positivo	a. Vero Positivo	b. Falso Positivo
Test Clinico Negativo	c. Falso Negativo	d. Vero Negativo

L'**accuratezza di un test** si ottiene dal rapporto tra le risposte corrette (veri positivi + veri negativi) ed il numero totale dei pazienti:

$$\text{Accuratezza Generale} = 100\% * (a + d)/(a + b + c + d)$$

Il valore di accuratezza diagnostica è influenzato dalla prevalenza di una patologia nella popolazione in un dato istante. Di conseguenza, per la valutazione della validità di un test clinico è preferibile utilizzare altre qualità psicometriche, quali ad esempio il valore predittivo positivo e negativo, la sensibilità, la specificità ed il rapporto di verosimiglianza.

### **VALORE PREDITTIVO POSITIVO (PPV) E VALORE PREDITTIVO NEGATIVO (NPV)**

Il valore predittivo positivo (PPV) è la capacità di un test di individuare i pazienti come malati rispetto alla positività del test diagnostico. Può essere anche definito come la percentuale di pazienti positivi al test in esame che sono realmente malati:

$$\text{PPV} = 100\% * a/(a + b)$$

Se il PPV è alto, un risultato positivo al test è fortemente indicativo della presenza della patologia.

Il valore predittivo negativo (NPV) è la capacità di un test di individuare i pazienti come sani rispetto alla negatività del test diagnostico. Rappresenta la percentuale di pazienti negativi al test in esame che sono realmente sani:

$$\text{NPV} = 100\% * d/(c + d)$$

### **SENSIBILITA'**

La sensibilità rappresenta la capacità o potenza di un test di individuare i pazienti come malati rispetto ad una condizione di malattia accertata da un esame diagnostico di riferimento (golden standard). In altri termini, è la percentuale dei pazienti realmente malati che risultano positivi al test in esame. L'utilità diagnostica dei test con alta sensibilità è l'esclusione di una particolare patologia. L'acronimo "**SnNout**" rappresenta una valida regola pratica secondo la quale, quando applicando un test altamente sensibile si ottiene un risultato negativo, possiamo escludere la patologia. Viceversa, se un test molto sensibile produce un risultato positivo, non è comunque possibile escludere la patologia, a causa della presenza dei falsi positivi (o errori di tipo alfa). La sensibilità di un test diagnostico è calcolata con la seguente formula matematica:

$$\text{Sensibilità} = 100\% * a/(a + c)$$

## SPECIFICITA'

La specificità è la capacità di un test di individuare i pazienti come sani rispetto ad una condizione di assenza di malattia accertata da un esame diagnostico di riferimento (golden standard). Può essere altresì definita come la percentuale dei pazienti realmente sani che risultano negativi al test in esame. I test diagnostici con alta specificità sono utili per accertare la presenza di una patologia. L'acronimo **"SpPin"** indica che un test con alta specificità, se positivo, conferma la patologia. Se invece un test altamente specifico è negativo, non è comunque possibile affermare con certezza l'assenza di malattia, a causa della presenza dei falsi negativi.

Un test sensibile sarà utilizzato per le patologie gravi e curabili, quindi quando è importante identificare un malato, mentre un test specifico sarà utilizzato quando il trattamento ha gravi effetti collaterali in modo da avere pochi falsi positivi oppure per confermare diagnosi già sospette.

La formula per calcolare la specificità di un test diagnostico è la seguente:

$$\text{Specificità} = 100\% * d/(b + d)$$

Un test ideale dovrebbe essere caratterizzato da sensibilità e specificità pari al 100%. Pochi test clinici dimostrano al contempo un'elevata sensibilità e specificità. Inoltre, la sensibilità e la specificità non ci danno indicazioni sulla variazione di probabilità che un paziente abbia una patologia, se il test risulta positivo o negativo. A tal fine ci viene in aiuto il **"Rapporto di verosimiglianza" o "Likelihood Ratio" (LR)**, che prende in considerazione anche questo aspetto.

**Tabella 5: Tabella di contingenza 2x2 con calcolo del PPV, NPV, sensibilità e specificità**

TEST DIAGNOSTICO	GOLDEN STANDARD POSITIVO	GOLDEN STANDARD NEGATIVO	
Test Clinico Positivo	a. Vero Positivo	b. Falso Positivo	PPV = a/(a+b)
Test Clinico Negativo	c. Falso Negativo	d. Vero Negativo	NPV = d/(c + d)
	Sensibilità = a/(a + c)	Specificità = d/(b + d)	

## **RAPPORTO DI VEROSIMIGLIANZA (LIKELIHOOD RATIO, LR)**

Il Likelihood ratio (LR) è il rapporto tra la frequenza con cui un certo valore del test si presenta tra i malati e la frequenza con cui lo stesso valore si presenta nei sani, o, meglio, nei NON affetti da quella malattia. Si ottiene dal rapporto fra sensibilità e specificità e permette di stabilire come si modificano le probabilità di una malattia da prima di eseguire il test (probabilità pre-test) a dopo averlo eseguito (probabilità post-test). In parole semplici, il LR rappresenta la capacità del risultato di un test di modificare la probabilità di un paziente di essere realmente malato (LR+) o realmente sano (LR-). Il LR consente di:

- ✓ valutare l'utilità di un esame diagnostico;
- ✓ scegliere l'esame più indicato per la specifica situazione del paziente.

Il likelihood ratio può essere positivo o negativo: un LR positivo aumenta la probabilità dell'esistenza di una patologia; un LR negativo aumenta la probabilità dell'assenza di una patologia. Il LR si calcola con le seguenti formule matematiche:

$$1) \text{ LR positivo} = \text{Sensibilità} / (1 - \text{Specificità})$$

$$2) \text{ LR negativo} = (1 - \text{Sensibilità}) / \text{Specificità}$$

Il LR può assumere diversi valori:

1. Se  $LR > 1$  il test è associato alla presenza di malattia
2. Se  $LR < 1$  indica che il test è associato all'assenza di malattia
3. Se  $LR = 1$  indica che il test non è informativo sulla malattia

La migliore visualizzazione grafica del significato del Rapporto di verosimiglianza è il **diagramma di Fagan** dove, interpolando i valori di probabilità pre test e il valore del Rapporto stesso, si ottiene il corrispondente valore di probabilità post test<sup>18</sup>.

Sono stati definiti dei criteri per interpretare i valori di LR, ottenuti dal rapporto tra sensibilità e specificità (tabella 6).

**Tabella 6: interpretazione dei rapporti di verosimiglianza**

LR Positivo	LR Negativo	Interpretazione del rapporto
Maggiore di 10	Minore di 0,1	Grande e decisiva variazione della probabilità
5-10	0,1-0,2	Moderata variazione della probabilità
2-5	0,2-0,5	Piccola, ma a volte importante variazione della probabilità
1-2	0,5-1,0	Piccola e raramente importante variazione della probabilità

## **CURVE DI ROC**

Le Curve di ROC rappresentano un modello quantitativo utilizzato per l'interpretazione del risultato di un test diagnostico. L'analisi ROC viene effettuata attraverso lo studio della funzione che, in un test quantitativo, lega la probabilità di ottenere un risultato vero positivo nella classe dei malati veri (ossia la sensibilità) alla probabilità di ottenere un risultato falso-positivo nella classe dei non-malati (ossia 1-specificità). La relazione tra i suddetti parametri è raffigurata attraverso una linea che si ottiene riportando, in un sistema di assi cartesiani e per ogni possibile valore di cut off, la proporzione di veri positivi in ordinata e la proporzione di falsi positivi in ascissa. Se il risultato del test è riportato su scala continua, si possono calcolare i valori di sensibilità e 1-specificità per ogni valore registrato.

La capacità discriminante di un test, ossia la sua attitudine a separare propriamente la popolazione in studio in "malati" e "sani" è proporzionale all'estensione dell'area sottesa alla curva ROC (AUC) ed equivale alla probabilità che il risultato di un test su un individuo estratto a caso dal gruppo dei malati sia superiore a quello di uno estratto a caso dal gruppo dei non-malati. L'area sottesa ad una curva ROC rappresenta pertanto un parametro fondamentale per la valutazione della performance di un test, in quanto costituisce una misura di accuratezza non dipendente dalla prevalenza ("pure accuracy").

Nel caso di un test perfetto, ossia che non restituisce alcun falso positivo né falso negativo (capacità discriminante = 100%), l'area sottesa alla curva ROC passa attraverso le coordinate (0;1) ed il suo valore corrisponde all'area dell'intero quadrato delimitato dai punti di coordinate (0,0), (0,1), (1,0) (1,1), che assume valore 1, corrispondendo ad una probabilità del 100% di una corretta classificazione. Si noti che, in tale caso, i valori predittivi non dipendono più dalla prevalenza. Al contrario, la ROC per un test assolutamente privo di valore informativo è rappresentata dalla diagonale che passa per l'origine, con un AUC pari a 0.5.

Per quanto riguarda l'interpretazione del valore dell'AUC, si può tenere presente la seguente classificazione della capacità discriminante:

- ✓ AUC=0.5 test non informativo
- ✓  $0.5 < \text{AUC} \leq 0.7$  test poco accurato
- ✓  $0.7 < \text{AUC} \leq 0.9$  test moderatamente accurato
- ✓  $0.9 < \text{AUC} < 1.0$  test altamente accurato
- ✓ AUC=1.0 test perfetto

Il vantaggio nell'utilizzo delle curve di ROC è rappresentato dalla possibilità di misurare l'informazione fornita da un test diagnostico sintetizzandola in un unico valore: ciò rende possibile confrontare tra di loro le accuratezze diagnostiche di test differenti per la stessa malattia in modo oggettivo<sup>19</sup>.

## **F1 SCORE**

L'F1 Score è un'altra misura della precisione ed accuratezza di un test diagnostico, che si ottiene dalla media armonica tra il Valore predittivo positivo e la sensibilità. E' espresso dalla seguente formula matematica:

$$\mathbf{F\ SCORE = 2*(PPV * Sensibilità)/(PPV + Sensibilità)}$$

L'F1 Score varia tra 0 e 1, tanto più si avvicina ad 1, tanto più il test diagnostico sarà accurato, viceversa se il valore è vicino allo zero, il test diagnostico sarà poco accurato<sup>20</sup>.

## 2.2. Risonanza Magnetica: Golden Standard

Le tecniche di imaging costituiscono parte integrante del processo di valutazione di una sindrome radicolare. Tra queste ricordiamo principalmente la Risonanza Magnetica Nucleare (RMN), la Tomografia Computerizzata (TC), la Mielografia, l'esame discografico e la valutazione chirurgica.

Per la valutazione dell'accuratezza diagnostica dei test di valutazione di una sindrome radicolare è necessario individuare un parametro di confronto "ottimale" o **Golden Standard**, inteso come il migliore e più accettato metodo per identificare una malattia o la condizione di interesse.

La Risonanza Magnetica Nucleare (RMN) è la tecnica di imaging più comunemente utilizzata per valutare le condizioni del tratto lombare della colonna vertebrale. Dalla risonanza magnetica è possibile esaminare bulging, protrusioni discali ed eventuali compressioni radicolari associate<sup>2</sup>.

Uno studio effettuato nel 2005 afferma che la sensibilità e la specificità della RMN nel rilevare la struttura nervosa compromessa della colonna vertebrale lombare è rispettivamente dell'83% e del 78%. E' stata rilevata, inoltre, un'affidabilità della RMN nella valutazione dell'ernia del disco lombare con un k pari a 0,74<sup>13</sup>.

Uno studio di **Modic** del 1986 esamina l'accuratezza ed il grado di concordanza diagnostica della RMN, TC, Mielografia tra di loro e comparati con il riscontro chirurgico, nella valutazione di 60 pazienti con sospetta ernia del disco o stenosi del canale lombare. Modic riscontrò un accordo dell'86,8% tra la RMN e la TC, dell'87,2% tra la RMN e la Mielografia, dell'82,6% tra la RMN ed il riscontro chirurgico per tipo e localizzazione della patologia, dell'83% tra la TC ed il riscontro chirurgico, del 71,8% tra la Mielografia ed il riscontro chirurgico. Combinando le diverse tecniche di imaging si ottennero i seguenti valori: 92,5% tra RMN e TC usate congiuntamente ed il riscontro chirurgico, 89,4% tra CT e Mielografia usate congiuntamente ed il riscontro chirurgico. I risultati indicano che la risonanza magnetica eseguita con una tecnica adeguata produce risultati diagnostici quantomeno equivalenti a quelli ottenuti con TC e Mielografia nella diagnosi dell'ernia del disco e della stenosi lombare. Risultati superiori potrebbero essere ottenuti utilizzando congiuntamente più tecniche di imaging, ma l'utilizzo della TC e della Mielografia presenta alcuni inconvenienti. Innanzitutto, con entrambe queste metodologie il paziente viene esposto ad una notevole quantità di radiazioni ionizzanti. Inoltre, la Mielografia è una procedura invasiva, quindi

meno gradita ai pazienti, spesso viene eseguita solo nei soggetti ricoverati in Ospedale e necessita del mezzo di contrasto intratecale<sup>21</sup>.

Uno studio di **Gibson** del 1986 effettua una comparazione tra la RMN e l'esame discografico nella valutazione della degenerazione discale, valutando in particolare 50 dischi intervertebrali. Lo studio mostra che la RMN produce immagini del nucleo polposo che sono comparabili a quelle prodotte con l'esame discografico. Pertanto, nella diagnosi della degenerazione discale, la RMN ha dimostrato di essere quantomeno tanto accurata quanto l'esame discografico, se non di più. Inoltre, la RMN presenta diversi vantaggi rispetto all'esame discografico: non è invasiva, è più sicura e gradita al paziente, non espone a radiazioni ionizzanti, è semplice da eseguire e da interpretare, pertanto ci sono meno rischi di errore da parte dell'osservatore. Per concludere quindi, la RMN, anche se utilizzata singolarmente, è la migliore tecnica di diagnosi dello stato di degenerazione discale, mentre all'esame discografico è relegato il ruolo di "*pain reproduction test*", da utilizzare in casi specificatamente selezionati<sup>22</sup>.

Diversi studi sostengono che la **TC** e la **RMN** sono pressoché sovrapponibili in termini di sensibilità, specificità e accuratezza diagnostica per porre diagnosi di ernia del disco e di stenosi del canale spinale. Nonostante ciò, la **RMN** è da considerarsi la tecnica di imaging elettiva per la diagnosi delle sindromi radicolari, in quanto non eroga radiazioni ionizzanti (temibili in particolare per le donne in età fertile), offre una valutazione anatomica realmente multiplanare (maggiore panoramicità rispetto alla TC), e visualizza meglio i tessuti molli e le alterazioni del tessuto osseo spongioso. Per le sue caratteristiche inoltre la RMN è in grado di evidenziare eventuali patologie intradurali (causa della sintomatologia o riscontro occasionale) che non sono rilevate dalla TC.

Considerando l'alta frequenza di ernie discali asintomatiche e la possibile regressione di ernie sintomatiche, è opportuno attendere almeno 4-6 settimane dall'insorgenza dei sintomi (dolore lombare, radicolopatia e limitazione funzionale) in assenza di *red flags*, prima di effettuare gli esami di diagnostica per immagini. TC o RMN sono invece da effettuare con urgenza in caso di sospetta malattia sistemica, neoplastica o infettiva, e di sospetta sindrome della cauda equina. TC o RMN sono invece raccomandate dopo 4-6 settimane di trattamento conservativo, in pazienti con sintomi o segni di compressione radicolare sufficientemente gravi da far considerare la possibilità dell'intervento chirurgico, anche in relazione alle preferenze del singolo paziente.

L'impiego dei mezzi di contrasto endovenosi con TC o RMN trova indicazione nella diagnostica dell'ernia del disco lombare solo in pazienti già sottoposti a intervento chirurgico e solo nei casi in cui lo studio senza mezzo di contrasto sia risultato non conclusivo.

La cosiddetta Mielo-TC (realizzata con l'introduzione di mezzo di contrasto intratecale, per rachicentesi) non è indicata nella diagnosi di ernia discale lombare per la sua invasività, i potenziali rischi e gli effetti collaterali. La Mielo-TC dovrebbe essere riservata solo a pazienti con patologia intradurale con controindicazione assoluta alla RMN.

A fronte dell'accresciuta offerta di prestazioni di diagnostica per immagini sempre più evolute, viene inoltre segnalata la necessità di migliorare l'appropriatezza del loro utilizzo, in modo da contenere i costi e il rischio radiologico. E' inoltre da tener presente che la sovradiagnosi e le indagini eseguite troppo precocemente potrebbero portare a indicazioni inappropriate al trattamento chirurgico, in casi destinati a risolversi spontaneamente<sup>23,24</sup>.

In alcuni studi sull'accuratezza diagnostica dei test per la valutazione delle sindromi radicolari viene utilizzato come *Golden Standard* il **riscontro chirurgico**. E' facile immaginare che l'accuratezza di tale procedura diagnostica è elevatissima; tuttavia presenta alcuni svantaggi. Il principale è che tale tecnica può essere applicata solo ad un campione ristretto ed altamente selettivo di pazienti che devono essere trattati con un intervento chirurgico e per i quali è molto più probabile che il test diagnostico abbia esito positivo. Inoltre, i pazienti con esito negativo ai test diagnostici preliminari all'intervento chirurgico ed alla diagnostica strumentale non potranno chiaramente essere valutati con il riscontro chirurgico<sup>5</sup>.

Possiamo pertanto affermare che la **Risonanza Magnetica Nucleare** rappresenta effettivamente il **GOLDEN STANDARD** per la diagnosi dell'ernia del disco, della stenosi lombare e delle conseguenti problematiche di intrappolamento della radice nervosa.

## 2.3. Test di valutazione della sindrome radicolare nel LBP acuto

### PREMESSE

I test per la valutazione della radicolopatia nel low back pain in fase acuta sono numerosi. In alcuni casi lo stesso test assume diverse denominazioni: ad esempio il prone knee bending test è identico al femoral stretch test ed alla manovra di Wassermann. Lo straight leg raising (SLR) test è identico alla manovra di Lasègue anche se, secondo alcuni autori, vi sono delle differenze. Lo stesso test può presentare delle varianti, che cercano di stressare in maniera più specifica alcune radici o componenti di uno stesso nervo. Non è però possibile avere test specificatamente selettivi per ogni nervo, in virtù della fitta rete di interconnessioni di cui è composto il sistema nervoso.

I sintomi del paziente vengono evocati attraverso differenti meccanismi: aumento della compressione radicolare, facendo assumere al paziente specifiche posizioni di chiusura del forame intervertebrale (principalmente le posizione di estensione del rachide) come nel Kemp's test e nell'hyperextension test oppure esercitando una pressione meccanica diretta a livello della radice nervosa, come nel caso del Bell Test. La maggior parte dei test agisce però attraverso un aumento della tensione a livello delle strutture neurali e per questo vengono denominati "test di neurotensione"; tra questi troviamo lo slump test, lo straight leg raising test normale ed incrociato, il prone knee bending test, ecc.

### TEST DI NEUROTENSIONE

Nell'esecuzione dei test di neurotensione si dovrà ricercare:

#### 1. Elicitazione del sintomo:

- ✓ grado di escursione in cui si manifestano i sintomi
- ✓ tipologia di sintomi
- ✓ copresenza di diversi sintomi
- ✓ tipologia di sintomi manifestati dal paziente alla fine dell'escursione

#### 2. Resistenza all'esecuzione del movimento

- ✓ punto in cui comincia la resistenza all'esecuzione movimento
- ✓ momento in cui la resistenza impedisce la prosecuzione del movimento
- ✓ qualità della resistenza (aumento graduale o improvviso)
- ✓ valutazione dei sintomi e delle resistenze dopo l'aggiunta delle diverse componenti del test.

## **CRITERI DI POSITIVITÀ DEI TEST DI NEUROTENSIONE**

Un test di neurotensione può essere considerato positivo se si verificano tali condizioni:

1. riproduce i sintomi del paziente
2. le risposte del test possono essere modificate dal movimento di parti distanti del corpo, per differenziarlo da problemi di origine muscoloscheletrica (test di differenziazione strutturale)
3. esistono differenze tra i due lati, riguardo alla risposta sintomatica, alla qualità della resistenza incontrata ed all'escursione articolare.

Per individuare una disfunzione di origine neurale è necessario effettuare:

- a. Esame soggettivo: le aree sintomatiche e la storia clinica del paziente possono indirizzare il clinico alle prime ipotesi diagnostiche
- b. Esame neurologico: deficit sensitivi e motori possono indicare il coinvolgimento di un tronco nervoso, di una radice o del midollo spinale.

## **Kemp's Test**

Il Kemp's test rappresenta uno dei due test, insieme allo Slump test, preliminari all'esame neurologico vero e proprio. Se risulta positivo, si procede con l'esame neurologico, viceversa quest'ultimo non viene effettuato. In casi di palese coinvolgimento radicolare, è possibile passare direttamente all'esame neurologico o addirittura rinviarne l'esecuzione, onde evitare di essere eccessivamente provocativi in fase acuta. Per quanto attiene alle modalità di esecuzione del Kemp's test, il paziente è seduto sul lettino con le ginocchia flesse a 90° e gli arti superiori incrociati. Il terapeuta si posiziona dietro al paziente e lo assiste bloccando con la mano caudale l'ala iliaca controlaterale al movimento da effettuare, mentre compie un movimento di estensione, flessione laterale e rotazione omolaterale del rachide dorso lombare. In tal modo, il sacco durale si sposta lateralmente, sul lato della rotazione. In caso di lesione discale laterale, il paziente lamenta un dolore caratterizzato da una componente radicolare omolaterale alla rotazione. Viceversa, in caso di lesione discale mediale, il paziente lamenta un dolore caratterizzato da una componente radicolare controlaterale alla rotazione<sup>26</sup>.

## Slump Test

Lo Slump test è indicato per la valutazione dell'estensibilità, della sensibilità e della reattività del sistema nervoso centrale, periferico e delle strutture vascolari ad esso connesse durante i movimenti di traslazione lungo un asse longitudinale. Questo test effettua una trazione sul sistema nervoso maggiore rispetto al suo movimento fisiologico. Durante il test si osservano tre parametri: provocazione del dolore, presenza di parestesie, restrizione di movimento. Tuttavia, la positività del test è data dall'evocazione della sintomatologia familiare al paziente e non dalla restrizione di movimento. Lo Slump test non è indicativo del livello della disfunzione. Il test si esaurisce nel momento in cui viene riprodotto il dolore del paziente. Si parte con la valutazione dell'arto non sintomatico, successivamente si passa ad esaminare quello sintomatico. Al termine viene effettuata una comparazione dei risultati nei due arti. Il test è composto da 7 fasi, in ciascuna delle quali aumenta progressivamente la tensione sul sistema nervoso. Ogni fase deve essere svolta prima attivamente dal paziente e poi passivamente dal fisioterapista.

Fasi del test:

1. Paziente seduto in posizione eretta con la mani dietro la schiena
2. Flessione toracica e lombare, senza flettere avanti il capo
3. Flessione della testa
4. Estensione del ginocchio attiva da parte del paziente, con ulteriore estensione passiva da parte del fisioterapista.
5. Flessione dorsale della caviglia, prima attiva e poi passiva
6. Estensione della colonna cervicale: se il dolore non cambia c'è un'alta possibilità che siano implicati i muscoli; se invece cambia c'è un'alta probabilità che sia presente una disfunzione di origine neurale
7. Verificare confrontando con l'arto controlaterale e poi con entrambe le gambe simultaneamente.

Per l'esecuzione dello Slump Test è necessaria una buona abilità manuale e capacità di interpretazione. Il test non va eseguito in caso di quadri sintomatologici irritabili. La componente di flessione potrebbe peggiorare sintomi correlati a quadri di origine discale, aumentando la pressione del disco o determinando una protrusione dello stesso contro il legamento longitudinale posteriore o la dura madre. Esistono alcune varianti dello Slump test, in posizione supina ed in long sitting.

Se lo Slump risulta positivo è necessario compiere l'esame neurologico. In caso di esito negativo dello Slump test non occorre compiere l'esame neurologico, poiché non c'è un coinvolgimento del sistema neurale. Se il dolore elicitato dal test non è quello familiare al paziente, la valutazione dipende dalla sede: se il dolore è lombare o toracico, si tratta di un dolore causato dalla distensione di strutture miofasciali; in tal caso il test viene pertanto considerato negativo<sup>26</sup>.

## **Straight Leg Raising (SLR)**

Storicamente lo straight leg raising test (SRL) è stato descritto da Lasègue, alunno di Forst, nel 1881 e successivamente dopo la sua morte da Lazarevic. Numerose variazioni e aggiunte si sono susseguite dopo la versione iniziale del test.

Lo *Straight Leg Raising test* (flessione della coscia a gamba tesa), nella versione di base, si effettua partendo con il paziente in posizione supina, senza l'utilizzo del cuscino, con i muscoli rilassati ed entrambe le ginocchia completamente estese. L'esaminatore, con una mano poggiata sul ginocchio del paziente, per mantenerlo in completa estensione e l'altra mano sotto il tallone, solleva lentamente l'arto inferiore sintomatico con la caviglia in posizione neutra, fino a quando compare il dolore (riferito come dolore "familiare" dal paziente) in regione lombosacrale, lungo l'arto inferiore o in entrambi. A questo punto l'esaminatore registra l'angolo di elevazione dell'arto inferiore in cui compaiono i sintomi. Viene chiesto al paziente di descrivere la distribuzione del dolore e/o la comparsa di parestesie<sup>25</sup>.

Affinché il test sia positivo per sindrome radicolare è necessario che sia riprodotto il dolore "familiare" del paziente, che deve comparire prima dei 60° di elevazione dell'arto e deve essere presente anche al di sotto del ginocchio. Il dolore evocato dal test è solitamente un dolore lancinante nella parte posteriore della coscia e spesso nella parte della gamba lungo il territorio di innervazione del nervo sciatico. I valori di sensibilità e specificità del SLR sono i seguenti:

Sensibilità:

- <60°: il test è positivo al di sotto di 60° nell'80% dei pazienti affetti da compressione radicolare;
- >60°: il test è positivo al di sopra di 60° nel 25% dei pazienti affetti da compressione radicolare;

Specificità:

- <60°: se il test è negativo, non c'è sindrome radicolare nel 40% dei casi;
- >60°: se il test è negativo, non c'è sindrome radicolare nel 90% dei casi.

L'SLR test stira principalmente le radici nervose da L4 ad S1. Un SLR test positivo, quindi, dovrebbe indicare una lesione da compressione delle radici da L4 a S1. Il test SLR si è dimostrato in grado di evidenziare anche lesioni a livelli più alti<sup>26</sup>.

Oltre alla versione descritta sopra, esistono alcune varianti dell'SLR che lo rendono maggiormente sensibile. Queste vengono scelte in base all'esito del test di base, durante il quale l'esaminatore deve registrare le eventuali aree di localizzazione dei sintomi di neurotensione avvertiti dal paziente. Conseguentemente, l'esaminatore decide quale variante è più indicata per confermare ovvero confutare le informazioni ottenute dal test di base. Dopo la sua esecuzione, si procede introducendo un movimento di un'articolazione lontana dal punto di insorgenza dei sintomi, evitando in tal modo di modificare la posizione delle strutture muscolo scheletriche dell'area sintomatica in esame. Si riesce così ad agire prevalentemente sulle strutture neurali; il cambiamento nella sintomatologia del paziente conferma che si tratta di una disfunzione di origine neurogenica. Per fare degli esempi pratici, se l'esito del test di base è un dolore prevalentemente a carico della gamba, si adotterà la variante di Neri o potranno essere aggiunti al test di base i movimenti di adduzione e rotazione interna dell'anca, se invece il sintomo è localizzato prevalentemente a livello del gluteo, verrà utilizzata la variante di Bragard<sup>26</sup>.

### ***Variante 1: Test di Bragard***

Prevede una dorsiflessione passiva del piede. Nel caso in cui il paziente avverte il dolore irradiato somministrando la versione "normale" dell'SLR, la variante di Bragard si effettua diminuendo leggermente la flessione a livello dell'anca e aggiungendo la flessione dorsale del piede. Questa manovra aumenta la tensione sul midollo spinale, sulla dura e sulle radici dei nervi lombosacrali rendendo più sensibile il test.

### ***Variante 2: Test di Neri***

Tale variante consiste nell'inserimento della flessione del capo attiva o assistita, diminuendo leggermente l'elevazione dell'arto quando compare il sintomo irradiato durante l'esecuzione dell'SLR.

### ***Variante 3: SLR con adduzione anca e/o rotazione interna dell'anca***

L'inserimento della rotazione interna dell'anca sollecita maggiormente la componente peroneale comune rispetto alla componente tibiale del nervo sciatico.

#### ***Variante 4: Discriminazione tra dolore dovuto allo stiramento muscolare e dolore dovuto all'irradiazione radicolare (Test di differenziazione strutturale)***

Per discriminare la componente muscoloscheletrica da quella neurale si esegue il test di differenziazione strutturale, che prevede le seguenti fasi:

- a. Appoggiare il piede del paziente sulla propria spalla mantenendo la gamba esaminata in posizione di SLR, mentre la mano craniale mantiene il ginocchio del paziente in completa estensione;
- b. Chiedere una flessione attiva della gamba, contraendo i muscoli ischio-tibiali.

Se si tratta di un problema muscolare, conseguentemente all'elicitazione del riflesso di rilassamento post-isometrico, dovrebbe essere possibile guadagnare alcuni gradi di escursione articolare ovvero il dolore dovrebbe diminuire. Se invece il dolore è radicolare la contrazione muscolare non fa cambiare nulla nella sintomatologia<sup>26</sup>.

#### **Crossed Straight Leg Raising Test (CSLR o XSLR)**

Il CSLR è un test molto specifico che si esegue in egual maniera all'SLR test e consiste nell'effettuare un sollevamento a ginocchio esteso dell'arto controlaterale a quello sintomatico. E' considerato positivo se compaiono i sintomi di tipo neurale nell'arto inferiore per il quale si ipotizza un problema di origine radicolare. Se il test evoca il sintomo del paziente sopra i 60° il risultato è positivo. Questo test ha un'altissima specificità per individuare segni di sofferenza radicolare (90%). Considerando che l'SLR ha un alta sensibilità ed il CSLR ha un'alta specificità, combinando i due test, otterremo un'altissima probabilità che il paziente sia affetto da sindrome radicolare<sup>26</sup>.

#### **Lasègue's Test**

Benché l'SLR sostanzialmente equivalga alla **manovra di Lasègue**, alcuni autori distinguono i due test. Tale manovra consiste nella flessione della coscia sul bacino a 90° a gamba flessa, seguita dall'estensione della gamba portata in linea con la coscia. Il test risulta positivo se il dolore irradiato al piede insorge o si aggrava solo con la prima manovra.

Di fatto, la manovra di Lasègue non prevede la misurazione dell'angolo, ma solo una risposta sì/no<sup>13</sup>.

Van Der Windt descrive il test di Lasègue come una variante dell'SLR, in cui l'elevazione dell'arto inferiore è ridotta di 5-10 gradi e, successivamente, viene effettuata una dorsiflessione del piede. Se con tale manovra viene evocato il dolore familiare al paziente, il test di Lasègue è considerato positivo<sup>5</sup>. In pratica, secondo Van Der Windt, la manovra di Lasègue corrisponderebbe alla variante di Bragard dell'SLR test.

## **Crossed Lasègue's Test**

Il **test di Lasègue incrociato o controlaterale** è identico a quello normale, precedentemente descritto, con la differenza che viene eseguito sull'arto inferiore sano e risveglia il dolore irradiato lungo il territorio di innervazione del nervo sciatico nell'arto inferiore controlaterale.

Entrambi i test (standard ed incrociato) rappresentano segni di sofferenza radicolare allo stiramento delle strutture neurali<sup>13</sup>.

## **Prone Knee Bending test (PKB)**

Il Prone Knee Bending test (PKB) esamina le menomazioni a livello delle radici nervose di L2, L3 o L4. Il paziente si posiziona in decubito prono. Il terapeuta posiziona la mano distale a livello del terzo distale della tibia e sul ginocchio, mentre la mano prossimale stabilizza il passaggio lombosacrale in modo da prevenire compensi da parte del paziente durante l'esecuzione del test. Il terapeuta effettua una flessione del ginocchio tra gli 80° ed i 100° per sviluppare il maggior allungamento della dura madre. Se in tal modo non viene evocato il dolore, estendere l'anca mantenendo sempre la flessione del ginocchio. Il test è positivo se c'è riproduzione del dolore unilaterale nell'area lombare, sul gluteo e/o sulla coscia anteriore. Anche questo test può essere reso più selettivo aggiungendo la flessione o l'estensione della testa. La sensibilità varia dall' 84% al 95%, mentre la specificità non è stata dimostrata<sup>26</sup>.

## **Femoral stretch test/Crossed femoral stretch test)**

Il femoral stretch test (FST) è identico al PKB e prevede anch'esso la posizione prona del paziente. La procedura di esecuzione è identica a quella del PKB. Il femoral stretch test incrociato (CFST) prevede la stessa procedura del FST, ma viene eseguito sull'arto inferiore sano, con l'obiettivo di elicitarne il dolore sempre sull'arto inferiore controlaterale a quello sano<sup>27</sup>.

## **Manovra di Wassermann**

La **manovra di Wasserman** si effettua con il paziente prono, in maniera quasi identica a quella descritta per il prone knee bending test. L'unica differenza, evidenziata da alcuni autori, riguarda il ginocchio, che dev'essere posizionato in massima flessione e non tra gli 80° ed i 100° di flessione. Spesso è positiva in caso di compressione sulle radici L2, L3 e L4, ma può essere positiva anche in caso di neuropatia femorale diabetica o di ematoma del muscolo psoas. In questi casi la manovra di Lasègue è in genere negativa poiché le radici L5 e S1 non sono coinvolte. La manovra induce tensione anche a livello delle strutture muscolari, in particolare del quadricipite (a causa del posizionamento in massima flessione del ginocchio) e ciò può determinare compensi posturali, quali ad esempio il sollevamento della natica e la rotazione omolaterale del bacino. I sintomi di tipo neurale si localizzano nella regione lombare alta e nella parte anteriore della coscia e dell'anca. Come per l'SLR test, anche per la manovra di Wassermann esistono delle varianti che sensibilizzano maggiormente alcune radici e/o componenti del nervo femorale. Ad esempio, l'aggiunta dell'estensione e l'adduzione dell'anca rendono il test maggiormente provocativo in caso di meralgia parestesica (tensione del nervo femoro-cutaneo laterale). L'estensione del rachide cervicale rappresenta un'ulteriore variante in grado di produrre maggiore sensibilizzazione delle strutture neurali<sup>27</sup>.

## **Bell Test**

Il Bell test si effettua con il paziente in posizione eretta. Il test consiste nella provocazione o esacerbazione del dolore radicolare attraverso una pressione eseguita con il pollice tra i processi spinosi di L4-L5 o tra L5-S1 o in corrispondenza dell'area paraspinale. Se la manovra riproduce solo dolore localizzato in regione lombare il risultato è negativo<sup>7</sup>.

## **Hyperextension Test**

Anche questo test prevede la posizione eretta del paziente. Il fisioterapista accompagna passivamente e gradualmente il tronco in massima estensione, assicurandosi che le ginocchia del paziente siano estese. Se la manovra provoca soltanto dolore lombare localizzato il risultato è negativo<sup>7</sup>.

## Finger To Floor Distance Test

Il paziente si posiziona su una piattaforma alta 20 cm in stazione eretta, senza scarpe ed a piedi uniti. L'esaminatore chiede al paziente di piegarsi in avanti fino al massimo grado di flessione possibile, mantenendo le ginocchia, le braccia e le dita completamente estese. La distanza verticale tra la punta del dito medio e la piattaforma viene misurata con un nastro flessibile e viene espressa in centimetri. Il test è positivo se il paziente non riesce a raggiungere la base della piattaforma con la punta del dito medio<sup>28</sup>.

Il finger to floor distance test ha un ottimo valore prognostico: la flessione del tronco fino a toccare il pavimento con la punta delle dita non solo stressa la radice nervosa ed il nervo sciatico, ma riduce anche la lordosi lombare ed accentua gli effetti meccanici dell'ernia discale sulla radice nervosa. Secondo Vroomen, questo test assume ancora maggior rilevanza se rapportato all'SLR, che non ha valore predittivo di compressione della radice nervosa, ma può indicare soltanto una sua tensione o irritazione. Vroomen inoltre indica, quale "*valore soglia*" del test, una distanza verticale tra la punta delle dita ed il pavimento di 25 cm, superata la quale c'è un'alta probabilità di presenza di compressione radicolare<sup>14</sup>.

Albeck, infine, sostiene che il finger to floor distance test è considerato suggestivo di ernia del disco se il paziente non riesce con la punta delle dita a superare il ginocchio durante la flessione del tronco<sup>29</sup>.

### 3. OBIETTIVI DELLO STUDIO

Questo studio si propone, in una prima fase, l'obiettivo di analizzare, attraverso la **Scala Quadas**, la qualità metodologica degli studi presenti in letteratura sull'accuratezza diagnostica dei principali test per la valutazione della sindrome radicolare nel LBP acuto. Successivamente, i test clinici relativi agli studi con valutazione Quadas più alta verranno valutati attraverso l'applicazione dell'**F1 Score**, al fine di individuare il test che presenta il miglior valore di accuratezza diagnostica.

La conoscenza dei valori di accuratezza diagnostica dei test per la radicolopatia lombare è di grande aiuto al fisioterapista nel ragionamento e nella pratica clinica, sia in fase di valutazione funzionale che di misurazione dell'outcome, durante ed alla fine del trattamento riabilitativo.

Una corretta identificazione della sindrome radicolare, basata su test di facile applicazione, ma al contempo accurati e validati in ambito scientifico, consente al fisioterapista di impostare efficacemente il trattamento riabilitativo, non più basato soltanto sulla risoluzione del sintomo, ma agendo direttamente anche sulla causa (bridge to therapy).

L'originalità del presente lavoro è rappresentata principalmente da due aspetti:

- 1) Determinare la qualità metodologica degli studi sull'accuratezza diagnostica dei test di valutazione della radicolopatia a livello lombare attraverso la **"SCALA QUADAS"**, strumento poco utilizzato in precedenza;
- 2) Applicare l'**F1 Score** agli studi che hanno ottenuto il punteggio più alto alla SCALA QUADAS, per individuare il test diagnostico più accurato.

## 4. MATERIALI E METODI

### 4.1 FONTI DI INFORMAZIONE

Per soddisfare gli obiettivi del presente studio è stata effettuata una revisione sistematica della letteratura scientifica riguardante l'accuratezza diagnostica dei test di valutazione della sofferenza radicolare nel low back pain acuto. La ricerca è stata condotta consultando banche dati elettroniche, in particolare Medline.

### 4.2 STRINGHE DI RICERCA

La prima stringa di ricerca è stata formulata con l'obiettivo di individuare studi riguardanti i più diffusi test clinici di valutazione dei segni di sofferenza radicolare a livello lombare. I termini chiave utilizzati nella ricerca sono stati i seguenti:

Search: "straight leg raise test" OR "straight leg raising test" OR "slr test" OR "crossed straight leg raising test" OR "lasègue" OR "lasègue's" OR "lasègue test" OR "lasègue's test" OR "lasègue sign" OR "lasègue's sign" OR "slump" OR "slump test" OR "kemp" OR "kemp's" OR "fingertip to floor test" OR "finger to floor distance" OR "bell" OR "bell's" OR "bell test" OR "X sign" OR "femoral stretch test" OR "femoral nerve stretch test" OR "wassermann" OR "wassermann's" OR "wassermann test" OR "wassermann's test" OR "bragard" OR "bragard's" OR "bragard test" OR "bragard's sign" OR "neri" OR "neri's" OR "knee flexion test" OR "valsalva maneuver" OR "valsalva's maneuver" OR "hyperextension test" OR "bowstring test" OR "bowstring sign".

Risultati: 52.611 items.

La seconda stringa di ricerca è stata formulata con l'obiettivo di individuare studi che analizzassero la sindrome di compressione radicolare all'arto inferiore ed il low back pain. I termini chiave utilizzati nella ricerca sono stati i seguenti:

Search: "nerve root compression" OR "radiculopathy" OR "sciatica" OR "nerve impingement" OR "low back pain".

Risultati: 27.548 items.

La terza stringa di ricerca è stata formulata con l'obiettivo di individuare studi riguardanti le qualità psicometriche dei test diagnostici, attraverso i seguenti termini di ricerca:

Search: "sensitivity" OR "specificity" OR "accuracy" OR "validity" OR "diagnosis" OR "differential" OR "reference standards".

Risultati: 3.649.488 items.

Infine, le precedenti stringhe di ricerca sono stati combinate, tramite l'utilizzo dell'operatore booleano "and", per ottenere studi riguardanti le qualità psicometriche dei test di valutazione dei sintomi di sofferenza radicolare nei pazienti affetti da low back pain acuto. In dettaglio, la stringa di ricerca finale è stata così formulata:

Search: (((("straight leg raise test" OR "straight leg raising test" OR "slr test" OR "crossed straight leg raising test" OR "lasègue" OR "lasègue's" OR "lasègue test" OR "lasègue's test" OR "lasègue sign" OR "lasègue's sign" OR "slump" OR "slump test" OR "kemp" OR "kemp's" OR "fingertip to floor test" OR "finger to floor distance" OR "bell" OR "bell's" OR "bell test" OR "X sign" OR "femoral stretch test" OR "femoral nerve stretch test" OR "wassermann" OR "wassermann's" OR "wassermann test" OR "wassermann's test" OR "bragard" OR "bragard's" OR "bragard test" OR "bragard's sign" OR "neri" OR "neri's" OR "knee flexion test" OR "valsalva maneuver" OR "valsalva's maneuver" OR "hyperextension test" OR "bowstring test" OR "bowstring sign"))) AND ("nerve root compression" OR "radiculopathy" OR "sciatica" OR "nerve impingement" OR "low back pain")) AND ("sensitivity" OR "specificity" OR "accuracy" OR "validity" OR "diagnosis" OR "differential" OR "reference standards").

Risultati: 152 items.

### **4.3 CRITERI DI INCLUSIONE ED ESCLUSIONE**

Le tipologie di studio oggetto della presente revisione sono state le seguenti: RCT, studi di coorte, studi caso-controllo, case report. Sono stati presi in considerazione studi in lingua inglese o italiana e con abstract consultabile, che avessero come oggetto la valutazione di pazienti con segni e sintomi di sofferenza radicolare diagnosticata attraverso Risonanza Magnetica, Tomografia Assiale Computerizzata, Mielografia o tramite riscontro diretto chirurgico.

I criteri di esclusione sono stati i seguenti:

- a. Articoli non in lingua inglese o italiana;
- b. Abstract non disponibile;
- c. Disegno dello studio non appropriato;
- d. Articoli di cui non è stato possibile ricavare il full text;

e. Articoli non rilevanti ai fini dell'indagine.

Una prima selezione è stata eseguita sulla base del titolo; successivamente è stata verificata la disponibilità dell'abstract ed analizzata la rilevanza dello studio in rapporto agli obiettivi della tesi. Un altro criterio di esclusione è stato il disegno dello studio. Le revisioni sistematiche sono state prese in considerazione per la parte relativa alle citazioni bibliografiche, utili per individuare ulteriori studi primari inerenti all'argomento trattato, eventualmente non emersi dalla ricerca.

Una seconda e più approfondita valutazione è stata effettuata dopo la lettura completa del full text degli articoli. Sono stati presi in considerazione solo gli articoli che rispondevano ai criteri di inclusione e che permettevano di soddisfare pienamente gli obiettivi dello studio. Gli studi validi alla fine della revisione, dopo lettura completa, sono risultati 15.

## 4.4 SCALA QUADAS

La Scala Quadas rappresenta un efficace strumento per la valutazione della qualità degli studi di accuratezza diagnostica. E' costituito da 14 domande, con 3 possibili risposte: "si", "no", "non chiaro."

**Tabella 7: The Quadas tool**

Item	Yes	No	Unclear
1. Was the spectrum of patients representative of the patients who will receive the test in practice?	( )	( )	( )
2. Were selection criteria clearly described?	( )	( )	( )
3. Is the reference standard likely to correctly classify the target condition?	( )	( )	( )
4. Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?	( )	( )	( )
5. Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?	( )	( )	( )
6. Did patients receive the same reference standard regardless of the index test result?	( )	( )	( )
7. Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?	( )	( )	( )
8. Was the execution of the index test described in sufficient detail to permit replication of the test?	( )	( )	( )
9. Was the execution of the reference standard described in sufficient detail to permit its replication?	( )	( )	( )
10. Were the index test results interpreted without knowledge of the results of the reference standard?	( )	( )	( )
11. Were the reference standard results interpreted without knowledge of the results of the index test?	( )	( )	( )
12. Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?	( )	( )	( )
13. Were uninterpretable/intermediate test results reported?	( )	( )	( )
14. Were withdrawals from the study explained?	( )	( )	( )

**Domanda n. 1: Le caratteristiche della popolazione dei partecipanti rappresentano quelle dei soggetti che nella pratica saranno soggetti al test?**

a) Cosa si intende con questa domanda: diverse caratteristiche demografiche e cliniche tra popolazioni diverse possono produrre misure di accuratezza diagnostica sensibilmente variabili. Si parla in questo caso di "spectrum bias". Questa domanda fa riferimento soprattutto alla generalizzabilità dei risultati piuttosto che all'eventualità che lo studio possa produrre risultati

affetti da bias. Se la popolazione dei soggetti sottoposti al test non è simile a quella sulla quale il test verrà applicato nella pratica, le stime di accuratezza diagnostica riportate avranno una applicabilità clinica limitata (cioè saranno poco generalizzabili). La popolazione dei soggetti riguarda non solo la rigorosità impiegata per la definizione della condizione di interesse ma anche le caratteristiche demografiche e la presenza di una diagnosi differenziale e/o di una comorbilità. È quindi importante che le valutazioni di un test diagnostico siano caratterizzate da una popolazione di soggetti appropriata al test oggetto della valutazione e da una descrizione chiara della popolazione realmente coinvolta nello studio.

*b) Situazioni in cui questo elemento non deve essere preso in considerazione:* questo elemento è rilevante per qualsiasi studio di accuratezza diagnostica e dovrebbe essere sempre incluso in uno strumento per la valutazione della qualità.

*c) Come valutare questo elemento:* la risposta deve essere “Sì” se si ritiene, in base alle informazioni riportate nello studio o in base a quelle reperibili dagli autori, che le caratteristiche dei soggetti inclusi nello studio siano rappresentative della popolazione di soggetti sul quale il test verrà utilizzato nella pratica. Il giudizio deve essere basato sia sul metodo di raccolta dei soggetti sia sulle caratteristiche di questi ultimi. La risposta a questo elemento deve essere “No” per quegli studi che reclutano un gruppo di soggetti sani come controllo e un gruppo che sicuramente sono affetti dalla patologia da diagnosticare. I revisori dovrebbero specificare nel protocollo della revisione quale popolazione di soggetti potrebbe essere accettabile, tenendo in considerazione fattori come gravità e prevalenza della patologia, età, sesso, ecc. Se si ritiene che la popolazione studiata non combacia con quanto è stato definito accettabile la risposta deve essere “No”. Se le informazioni disponibili sono insufficienti per dare un giudizio, la risposta deve essere “Non chiaro”.

## **Domanda n. 2: I criteri di selezione sono descritti chiaramente?**

*a) Cosa si intende con questa domanda:* con questa domanda si intende capire se lo studio contiene una definizione chiara dei criteri usati per l’inclusione e l’esclusione dei soggetti all’interno dello studio.

*b) Situazioni in cui questo elemento non deve essere preso in considerazione:* questo elemento è rilevante per qualsiasi studio di accuratezza diagnostica e dovrebbe essere sempre incluso in uno strumento per la valutazione della qualità.

*c) Come valutare questo elemento:* la risposta deve essere “Sì” se si ritiene che siano state fornite tutte le informazioni in merito a come i soggetti sono stati selezionati. La risposta deve essere “No” se i criteri di selezione non sono chiaramente riportati. Se i criteri sono riportati solo

parzialmente e si ritiene di non avere abbastanza informazioni per rispondere “Sì”, la risposta deve essere “Non chiaro”.

**Domanda n. 3: Lo standard di riferimento è adatto a classificare correttamente la condizione di interesse?**

*a) Cosa si intende con questa domanda:* lo standard di riferimento (o Golden Standard) è il metodo utilizzato per determinare l’assenza o la presenza della condizione d’interesse. Per valutare l’accuratezza diagnostica del test indice (test sotto esame), il risultato di quest’ultimo viene comparato con il risultato dello standard di riferimento; successivamente sarà possibile calcolare gli indicatori dell’accuratezza diagnostica. Lo standard di riferimento è quindi un determinante importante dell’accuratezza diagnostica di un test. Le stime sulle performance di un test sono basate sull’assunzione che lo standard di riferimento possieda il 100% di sensibilità e il 100% di specificità. Se si verificano delle incongruenze tra i risultati dello standard di riferimento e quelli del test indice, si assume che questi ultimi valori non siano corretti. Quindi, secondo un punto di vista teorico, la scelta di uno standard di riferimento appropriato è molto importante.

*b) Situazioni in cui questo elemento non deve essere preso in considerazione:* questo elemento è rilevante per qualsiasi studio di accuratezza diagnostica e dovrebbe essere sempre incluso in uno strumento per la valutazione della qualità.

*d) Come valutare questo elemento:* La risposta deve essere “Sì” se si ritiene che lo standard di riferimento sia in grado di classificare correttamente la condizione di interesse o se è ritenuto il miglior metodo disponibile per farlo. Dare un giudizio sull’accuratezza dello standard di riferimento può non essere cosa facile. È necessaria infatti la conoscenza approfondita dell’area di studio per sapere se un determinato test può essere assunto come standard di riferimento o se una combinazione di test può essere più accurata. La risposta deve essere “No” se si ritiene che lo standard di riferimento non sia in grado di classificare correttamente la condizione di interesse. Se le informazioni disponibili sono insufficienti per dare un giudizio, la risposta deve essere “Non chiaro”.

**Domanda n. 4: L’intervallo temporale tra lo standard di riferimento e il test indice è abbastanza breve per essere ragionevolmente sicuri che la condizione di interesse non subisca variazioni tra i due test?**

*a) Cosa si intende con questa domanda:* Idealmente, i risultati del test indice e dello standard di riferimento vengono raccolti dallo stesso soggetto nello stesso istante. Se questo non è possibile e c’è un ritardo tra i due campionamenti, possono verificarsi delle classificazioni errate dovute al recesso spontaneo o alla progressione dello stato patologico. Si parla allora di “disease

*progression*” bias. La lunghezza del periodo di tempo capace di causare tale bias varierà in base alla patologia ed al relativo stadio di avanzamento. Ad esempio, un ritardo di pochi giorni difficilmente sarà un problema in caso di una condizione cronica, tuttavia, per molte malattie infettive lo stesso ritardo tra il momento del test indice e quello dello standard di riferimento può essere molto importante. Questo tipo di bias può verificarsi anche in quelle condizioni croniche nelle quali lo standard di riferimento richiede un follow-up clinico di diversi anni.

*b) Situazioni in cui questo elemento non deve essere preso in considerazione:* questo elemento potrebbe essere applicato in molte situazioni.

*c) Come valutare questo elemento:* per condizioni con progresso rapido anche un ritardo di qualche giorno può essere importante. Per tali condizioni, se il ritardo tra il test indice e lo standard di riferimento è dell’ordine delle ore o del singolo giorno, la risposta deve essere “Sì”. Nelle patologie croniche invece sarà difficile osservare variazioni nell’arco di una settimana o di un mese. Quindi in questi casi, anche un ritardo lungo potrà portare ad un “Sì” nella risposta. In questo elemento è necessario dare dei giudizi riguardo alla definizione di “abbastanza breve”. La lunghezza dell’intervallo dovrà essere attentamente rapportata allo specifico segmento di interesse. La risposta deve essere “No” se si ritiene che l’intervallo di tempo tra il test indice e il test di riferimento sia sufficientemente lungo da comportare delle variazioni nella condizione patologica del paziente. Se le informazioni disponibili non sono insufficienti per dare un giudizio, la risposta deve essere “Non chiaro”.

**Domanda n. 5: L’intero campione o una selezione casuale del campione sono state sottoposte a verifica attraverso uno standard di riferimento della diagnosi?**

*a) Cosa si intende con questa domanda:* quando nessun gruppo dello studio è stato sottoposto alla conferma della diagnosi attraverso lo standard di riferimento può verificarsi il “*partial verification*” bias (noto anche come *workup bias*, *selection bias* o *sequential ordering bias*). Se il risultato del test indice influenza la decisione di effettuare lo standard di riferimento, le stime delle performance del test possono essere affette da bias. Se su una selezione casuale dei pazienti viene effettuato lo standard di riferimento, le performance globali del test rimangono teoricamente invariate. In molti casi, tale selezione non è casuale è questo comporta un bias sulle stime dell’accuratezza diagnostica globale.

*b) Situazioni in cui questo elemento non deve essere preso in considerazione:* generalmente il *partial verification bias* si presenta soltanto negli studi diagnostici di coorte nei quali i pazienti vengono sottoposti al test indice e successivamente allo standard di riferimento. Nelle situazioni in cui lo standard di riferimento viene effettuato prima del test indice, occorre decidere a priori se

esiste la possibilità che si verifichi il partial verification bias . Questo elemento può essere legato alla qualità finale della revisione. Stando a quando detto, si può rispondere “Sì” oppure si può rimuovere questo elemento.

c) *Come valutare questo elemento*: la risposta deve essere “Sì” se risulta chiaramente che tutti i pazienti che sono stati sottoposti al test indice, o una selezione casuale di essi, ha ricevuto una conferma della diagnosi attraverso lo standard di riferimento. La risposta deve essere “Sì” anche se lo standard di riferimento non è lo stesso per tutti i pazienti. La risposta deve essere “No” se alcuni dei pazienti sottoposti al test indice non hanno ricevuto conferma della diagnosi oppure se la selezione che ha ricevuto conferma attraverso lo standard di riferimento non è casuale. La risposta deve essere “Non chiaro” se queste informazioni non sono riportate nello studio.

**Domanda n. 6: I partecipanti sono stati sottoposti allo stesso standard di riferimento indifferentemente dall’esito del test indice?**

a) *Cosa si intende con questa domanda*: quando alcuni dei risultati ottenuti con il test indice sono verificati attraverso diversi standard di riferimento, si parla di “*differential verification*” bias. Questo può essere considerato un problema se i vari standard di riferimento differiscono tra di loro per la definizione della condizione di interesse. Questo accade di solito quando, i pazienti trovati positivi al test indice vengono sottoposti a verifica con un metodo molto più accurato (spesso invasivo) rispetto a quelli che vengono trovati negativi. La correlazione tra un certo risultato del test (negativo) e la sua verifica attraverso uno standard di riferimento meno accurato può influenzare la misura dell’accuratezza del test con un bias simile al partial verification bias, ma meno pesante.

b) *Situazioni in cui questo elemento non deve essere preso in considerazione*: il differential verification bias può verificarsi in tutti i tipi di studi sull’accuratezza diagnostica.

c) *Come valutare questo elemento*: la risposta deve essere “Sì” se è chiaro che i pazienti che hanno ricevuto conferma della loro condizione sono stati sottoposti allo stesso standard di riferimento. La risposta deve essere “No” se alcuni pazienti sono stati sottoposti a verifica attraverso standard di riferimento diversi. Se queste informazioni non sono riportate nello studio, la risposta deve essere “Non chiaro”.

**Domanda n. 7: Lo standard di riferimento era indipendente dal test indice (cioè, il test indice non è parte dello standard di riferimento)?**

a) *Cosa si intende con questa domanda*: quando il risultato del test indice è usato per stabilire la diagnosi finale, si parla di “*incorporation bias*”. Questo tipo di bias porta a sovrastimare le misure di accuratezza diagnostica del test indice, poiché aumenta le congruenze tra i risultati del test

indice e quelle dello standard di riferimento. È importante sottolineare che la sola conoscenza dei risultati del test indice non significa che automaticamente tali risultati sono incorporati nello standard di riferimento. Ad esempio, uno studio che considera la MRI (risonanza magnetica per immagini) per la diagnosi della sclerosi multipla può avere uno standard di riferimento composto da: follow-up clinico, analisi CSF (analisi del fluido cerebro-spinale) e MRI. In questo caso il test indice è parte dello standard di riferimento. Se lo stesso studio utilizza il follow-up clinico come standard di riferimento e i risultati della MRI erano noti quando è stata fatta la diagnosi clinica ma non erano esplicitamente definiti come parte dello standard di riferimento, allora si può dire che il test indice non è parte dello standard di riferimento.

*b) Situazioni in cui questo elemento non deve essere preso in considerazione:* questo elemento può essere applicato solo quando per la verifica dello stato della patologia viene utilizzato uno standard di riferimento composto. In tali casi è essenziale definire precisamente come viene verificato lo stato della patologia e occorre dichiarare quali test sono parte dello standard di riferimento. Per gli studi che usano un singolo standard di riferimento questo elemento non è rilevante e può essere marcato come “Sì” o può essere rimosso.

*c) Come valutare questo elemento:* la risposta deve essere “Sì” se dallo studio si evince chiaramente che il test indice non è parte dello standard di riferimento. La risposta deve essere “No” se il test indice è parte dello standard di riferimento. Se queste informazioni non sono riportate nello studio, la risposta deve essere “Non chiaro”.

#### **Domande n. 8 e 9:**

**L’esecuzione del test indice è stata descritta con sufficiente dettaglio da permetterne la ripetizione?**

**L’esecuzione dello standard di riferimento è stata descritta con sufficiente dettaglio da permetterne la ripetizione?**

*a) Cosa si intende con queste domande:* una descrizione sufficientemente chiara dell’esecuzione del test indice e del test di riferimento è necessaria per due motivi. Primo, alcune variazioni nella misura dell’accuratezza diagnostica possono essere dovute a delle differenze nell’esecuzione del test indice o dello standard di riferimento. Secondo, una descrizione chiara e dettagliata (o una citazione) può essere necessaria per ripetere i test in un contesto diverso. Se i test sono condotti in modo diverso è facile aspettarsi differenze nelle performance. Ovviamente tutto dipende dal tipo di test sotto indagine.

*b) Situazioni in cui questi elementi non devono essere presi in considerazione:* questi elementi potrebbero essere applicati in molte situazioni.

c) *Come valutare questi elementi*: la risposta deve essere “Sì” se lo studio contiene dettagli o citazioni sufficienti per consentire la replica del test indice e dello standard di riferimento. La risposta deve essere “No” per tutti gli altri casi. Se i dettagli sono riportati parzialmente ed esiste l’evidenza che non siano state fornite tutte le informazioni necessarie per replicare i test, la risposta deve essere “Non chiaro”.

**Domande n. 10 e 11:**

**I risultati del test indice sono stati interpretati senza essere a conoscenza dei risultati dello standard di riferimento?**

**I risultati dello standard di riferimento sono stati interpretati senza essere a conoscenza dei risultati del test indice?**

a) *Cosa si intende con queste domande*: questo elemento è simile al blinding (accecamiento) degli studi di intervento. L’interpretazione dei risultati del test indice può essere influenzata dalla conoscenza dei risultati dello standard di riferimento e viceversa. Si parla in questo caso di “*review bias*”. Il grado di influenzabilità è chiaramente legato al grado di soggettività che caratterizza l’interpretazione del risultato del test. Più l’interpretazione è soggettiva e più probabilmente colui che interpreta i risultati del test indice sarà influenzato dai risultati del test di riferimento e viceversa. È quindi importante considerare la specifica area di interesse per determinare a priori se l’interpretazione dei risultati dei test possa subire delle influenze dovute alla conoscenza dei risultati di altri test.

b) *Situazioni in cui questi elementi non devono essere presi in considerazione*: se il test indice viene sempre effettuato per primo, allora l’interpretazione dei risultati generalmente sarà effettuata senza la conoscenza dei risultati dello standard di riferimento. In modo analogo, se lo standard di riferimento è sempre effettuato per primo (ad esempio in uno studio diagnostico di tipo case-control), allora l’interpretazione dei risultati generalmente sarà effettuata senza la conoscenza dei risultati del test indice. Tuttavia, se i risultati sono interpretati in un secondo momento, dopo che sia stato effettuato sia il test indice che lo standard di riferimento, è molto importante che lo studio chiarisca se l’interpretazione di ciascun gruppo di risultati sia stata fatta con metodo blind. Nelle situazioni in cui il review bias non appare importante si può rispondere “Sì” a entrambe le domande o rimuoverle entrambe. Se i test sono assolutamente oggettivi nella loro interpretazione allora tale interpretazione non può essere affetta da review bias. Altra situazione in cui questi elementi appaiono irrilevanti è quella che prevede il coinvolgimento di un laboratorio indipendente per l’interpretazione dei risultati poiché è molto improbabile che gli operatori siano a conoscenza dei risultati dei test che non sono stati effettuati nel loro laboratorio.

c) *Come valutare questi elementi:* la risposta deve essere “Sì” se lo studio spiega chiaramente che i risultati dei test (test indice e standard di riferimento) sono stati interpretati con metodo blind. La risposta deve essere “No” se tale spiegazione non è fornita. Se questa informazione non è riportata nello studio, la risposta deve essere “Non chiaro”.

**Domanda n. 12: I dati clinici utilizzati durante lo studio sono disponibili normalmente nella pratica clinica quotidiana?**

a) *Cosa si intende con questa domanda:* la disponibilità di dati clinici durante l’interpretazione dei risultati del test può influire sulle stime della performance del test stesso. In questo contesto, il termine “dati clinici” assume un significato ampio, arrivando a comprendere dati ottenuti dall’osservazione diretta del paziente come età, sesso, sintomi. La conoscenza di tali fattori può influenzare i risultati di un test diagnostico se quest’ultimo è caratterizzato da una componente interpretativa. Se i dati clinici sono disponibili al momento dell’interpretazione del test nella pratica, allora sono disponibili anche quando il test è in fase di valutazione. Se tuttavia, il test indice intende sostituire uno o più altri test clinici, allora i dati clinici non dovranno essere disponibili o dovranno essere disponibili per tutti i test in esame. È quindi importante definire quali informazioni devono essere disponibili quando i risultati del test saranno interpretati nella pratica.

b) *Situazioni in cui questo elemento non deve essere preso in considerazione:* questo elemento può tranquillamente essere omesso se l’interpretazione del test indice è completamente automatica e non prevede nessuna interpretazione.

c) *Come valutare questo elemento:* la risposta deve essere “Sì” se i dati clinici che sono disponibili al momento dell’interpretazione dei risultati nella pratica erano disponibili al momento dell’interpretazione del test indice nello studio. In modo analogo, la risposta deve essere “Sì” se i dati clinici al momento dell’interpretazione dei risultati non sono disponibili nella pratica e nemmeno al momento dell’interpretazione del test indice nello studio. La risposta deve essere “No” in tutti gli altri casi. Se queste informazioni non sono riportate nello studio, la risposta deve essere “Non chiaro”.

**Domanda n. 13: Sono riportati risultati del test intermedi/non interpretabili?**

a) *Cosa si intende con questa domanda:* un test diagnostico può produrre risultati indeterminati/intermedi/non interpretabili con frequenza variabile in relazione al tipo di test. Questi problemi spesso non sono riportati negli studi di accuratezza diagnostica, all’interno dei quali, i risultati non interpretabili vengono semplicemente rimossi dall’analisi. Tale rimozione può influenzare la valutazione delle caratteristiche del test se esiste una correlazione tra i risultati non

interpretabili e la condizione vera della patologia. Se i risultati non interpretabili si presentano casualmente e non sono legati alla diagnosi reale, allora, in teoria, non avranno alcun effetto sulla performance del test. Qualunque sia la causa dei risultati non interpretabili, è importante riportare anche essi nello studio.

*b) Situazioni in cui questo elemento non deve essere preso in considerazione:* questo elemento è rilevante per qualsiasi studio di accuratezza diagnostica e dovrebbe sempre essere incluso in uno strumento per la valutazione della qualità.

*c) Come valutare questo elemento:* la risposta deve essere “Sì” se risulta chiaro che tutti i risultati dei test (anche quelli indeterminati/intermedi/non interpretabili) siano stati riportati. La risposta deve essere “Sì” anche se gli autori non riportano risultati indeterminati/intermedi/non interpretabili tuttavia riportano i risultati per tutti i partecipanti inclusi nello studio. La risposta deve essere “No” se si ritiene che non tutti i risultati siano stati riportati nello studio. Se non è possibile capire se tutti i risultati sono stati riportati, la risposta deve essere “Non Chiaro”.

#### **Domanda n. 14: Sono stati spiegati i ritiri dallo studio?**

*a) Cosa si intende con questa domanda:* Può verificarsi che alcuni pazienti vengano ritirati dallo studio (bias di attrito) prima dei risultati del test indice e/o di quelli dello standard di riferimento. Se i soggetti ritirati differiscono sensibilmente, per qualsiasi ragione, da quelli rimasti all’interno del gruppo di studio, questo può influenzare la stima delle performance del test.

*b) Situazioni in cui questo elemento non deve essere preso in considerazione:* questo elemento è rilevante per qualsiasi studio di accuratezza diagnostica e dovrebbe sempre essere incluso in uno strumento per la valutazione della qualità.

*c) Come valutare questo elemento:* la risposta deve essere “Sì” se si evince chiaramente cosa è successo a tutti i pazienti entrati nello studio, ad esempio con l’aiuto di un diagramma di flusso dei partecipanti. La risposta deve essere “Sì” anche se gli autori non riportano nessun ritiro dallo studio tuttavia sono disponibili i dati relativi a tutti i partecipanti coinvolti nello studio. La risposta deve essere “No” se si nota che alcuni pazienti non sono arrivati al termine dello studio (ad esempio, non sono stati sottoposti ad entrambi i test, test indice e standard di riferimento). Se non è noto se tutti i pazienti entrati nello studio sono giunti al termine, la risposta dev’essere “Non Chiaro”<sup>30</sup>.

Tutti gli studi considerati validi ai fini della revisione sono stati valutati con la Scala Quadras. A parità di punteggio, si è deciso di dare maggior peso ai “No” rispetto ai “Non Chiaro” e di privilegiare pertanto gli studi con minor numero di “No”.

Gli studi con punteggio Quadas maggiore sono stati successivamente valutati con l'**F1 Score**, ulteriore misura per la valutazione della precisione ed accuratezza di un test, con l'obiettivo di individuare il test con il maggior livello di accuratezza diagnostica per la valutazione dei segni e sintomi di sofferenza radicolare nel low back pain acuto.

## 5. RISULTATI

### 5.1: Diagramma di flusso: selezione degli articoli

Inserendo i termini chiave su Pubmed sono stati individuati 152 studi, a cui vanno aggiunti altri 23 studi identificati nelle bibliografie delle revisioni sistematiche presenti nella prima ricerca, per un totale di **175** studi.

In seguito ad una prima selezione che ha preso in considerazione il titolo, l'abstract, il disegno dello studio e la lingua sono stati esclusi **128** studi.

La ricerca della versione completa (full text) degli articoli selezionati ha comportato l'esclusione di **25** studi, in quanto non reperibili.

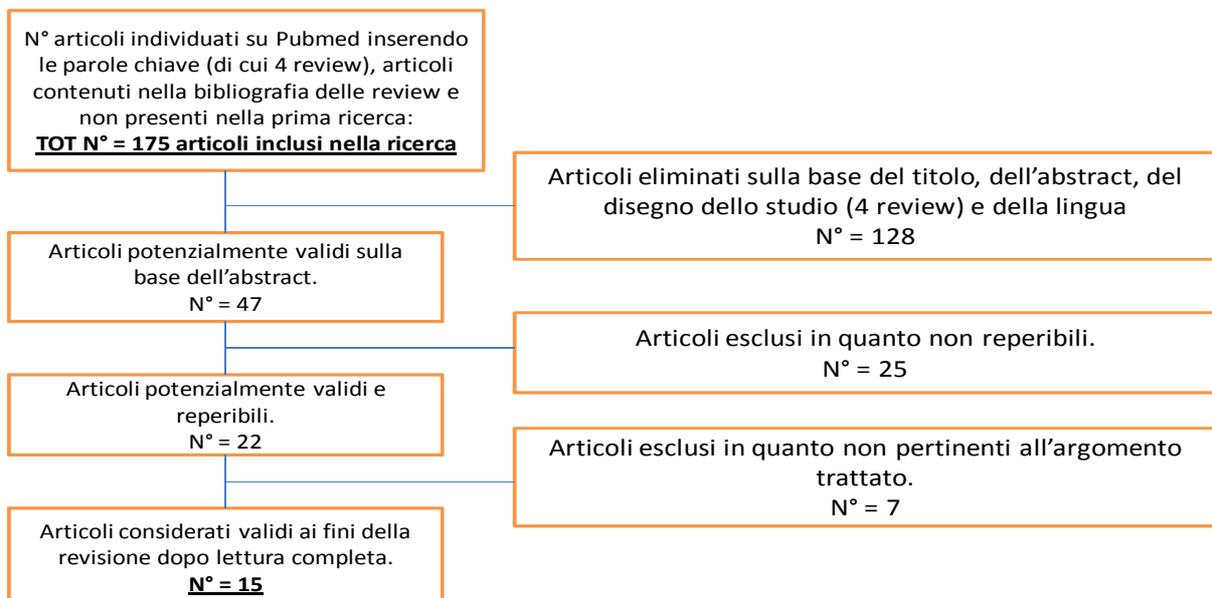
La seconda selezione è stata effettuata in seguito alla lettura completa di ciascun articolo. Sono stati esclusi dalla revisione **7** studi in quanto non soddisfacevano gli obiettivi dello studio.

La ricerca si conclude con **15** articoli definitivamente inclusi nella revisione.

I passaggi dei metodi di selezione sono di seguito schematizzati in un diagramma di flusso.

Tabella 8: diagramma di flusso: selezione degli articoli

### DIAGRAMMA DI FLUSSO: SELEZIONE DEGLI ARTICOLI



## 5.2 Analisi descrittiva degli studi oggetto della revisione

Vediamo nel dettaglio le caratteristiche peculiari degli studi oggetto della presente revisione, con l'obiettivo di evidenziare contributi e limiti degli stessi, in relazione all'accuratezza diagnostica dei test clinici per la valutazione dei segni e sintomi di sofferenza radicolare a livello lombare.

**Capra F., Vanti C., Donati R., Tombetti S., O'Reilly C., Pillastrini P. – Validity of the straight leg raise test for patients with sciatic pain with or without lumbar pain using magnetic resonance imaging results as a reference standard<sup>31</sup>.**

Lo scopo di questo studio è valutare la validità dell'SLR, utilizzando la risonanza magnetica come golden standard, in un gruppo di 2.352 pazienti con ernia del disco lombare a livello di L4-L5 e L5-S1 e sciatalgia. L'autore ha analizzato la relazione tra accuratezza diagnostica dell'SLR, classi di età (da 16 a 85 anni) e grado di protrusione discale. L'accuratezza diagnostica è stata espressa attraverso i valori di sensibilità, specificità, valore predittivo positivo e negativo, odds ratio, rapporto di verosimiglianza, curva ROC. I valori riscontrati sono stati i seguenti: sensibilità 0,36, specificità 0,74, valore predittivo positivo 0,69, valore predittivo negativo 0,52, LR+ 1,38, LR- 0,87, odds ratio 1,59, ROC 0,596 con una riduzione pari a 0,730 nella classe di età 16-25 anni ed a 0,515 nella classe di età 76-85 anni. *Tali risultati evidenziano un basso livello di accuratezza diagnostica dell'SLR nella valutazione dei pazienti con segni e sintomi di radicolopatia, in rapporto ai risultati ottenuti con la risonanza magnetica. Inoltre, il potere diagnostico dell'SLR sembra essere inversamente proporzionale all'età.*

**Rabin A., Gerszten P.C., Karausky P., Bunker C.H., Potter D.M., Welch W.C. – The sensitivity of the Seated Straight-Leg Raise Test Compared With the Supine Straight-Leg Raise Test in Patients Presenting With Magnetic Resonance Imaging Evidence of Lumbar Nerve Compression<sup>2</sup>.**

Lo studio di Rabin del 2007 pone a confronto due versioni dello Straight leg raise (SLR) test, la prima in posizione supina, l'altra in posizione seduta. Quest'ultima presenta due vantaggi: riduce i tempi di realizzazione ed è più confortevole per il paziente. Tuttavia, in pazienti con evidenza di compressione radicolare alla risonanza magnetica, l'SLR da supino possiede un livello di sensibilità pari a 0,67, mentre lo stesso test da seduto ha una sensibilità pari a 0,41. Pertanto, *il test da supino riflette in maniera più accurata i riscontri radiologici di compressione radicolare a livello lombare.* Rabin non fornisce indicazioni in merito ai livelli di specificità delle due versioni del test.

Lo studio evidenzia inoltre che la sensibilità dell'SLR da supino è più bassa di quella mediamente rilevata, con un valore di 0,67 rispetto ad un valore medio pari a 0,91. Questa differenza è dovuta al fatto che altri studi hanno utilizzato, come golden standard, l'esplorazione chirurgica invece della risonanza magnetica. Inoltre, il campione preso in considerazione in questi studi era rappresentato da pazienti sottoposti ad intervento chirurgico, mentre lo studio di Rabin ha analizzato 71 pazienti con segni e sintomi di sofferenza radicolare. La minore sensibilità potrebbe anche essere dovuta ai criteri di positività adottati (ad esempio, chiara riproduzione del dolore irradiato al di sotto del ginocchio), differenti nei diversi studi. Infine, l'utilizzo di farmaci antinfiammatori o di steroidi da parte di alcuni pazienti prima della somministrazione del test potrebbe aver ridotto l'irritazione della radice nervosa, con conseguente diminuzione del livello di sensibilità dell'SLR. Nonostante ciò, questo test resta uno dei test basilari per un buon esame fisico nei casi di sospetta sindrome radicolare a livello lombare.

***Xin S.Q., Zhang Q.Z., Fan D.H. – Significance of the straight-leg-raising test in the diagnosis and clinical evaluation of lower lumbar intervertebral-disc protrusion<sup>25</sup>.***

Questo autore nel suo studio analizza 113 pazienti affetti da protrusione discale a livello lombare e mostra una forte interconnessione tra il riscontro chirurgico (utilizzato in questo caso come golden standard) ed i risultati dell'SLR test in termini di sede della protrusione discale e localizzazione del dolore. Inoltre, la limitazione di ROM, evidenziata durante l'esecuzione dell'SLR, ha una relazione diretta con la sede, il grado di protrusione discale e dell'eventuale compressione radicolare. *Durante la somministrazione dell'SLR, una protrusione centrale tende a provocare dolore a livello lombare, una protrusione laterale causa dolore irradiato agli arti inferiori, una protrusione intermedia causa entrambi i tipi di dolori. Partendo da tali premesse, la distribuzione del dolore durante l'SLR ha permesso di determinare, in maniera accurata, la localizzazione della lesione nell'88,5% dei pazienti esaminati.* In tali pazienti, il dolore veniva evocato tra i 40 ed i 65 gradi di flessione dell'anca a ginocchio esteso.

***Kosteljanetz M., Espersen J.O., Halaburt H., Miletic T. – Predictive value of clinical and surgical findings in patients with lumbago-sciatica. A prospective study (Part I)<sup>35</sup>.***

In questo studio vengono analizzati 100 pazienti con dolore lombare e sciatica, possibili candidati all'intervento chirurgico. A tali pazienti viene somministrato l'SLR test, con lo scopo di distinguere i pazienti con sintomi di tipo radicolare che dovranno essere sottoposti ad intervento chirurgico da quelli per i quali è sufficiente un trattamento conservativo. Lo studio si propone altresì di valutare

il valore predittivo dell'SLR, in riferimento all'individuazione del livello coinvolto e del tipo di lesione, responsabili della compressione radicolare. Il golden standard in questo caso è rappresentato principalmente dal riscontro chirurgico, in secondo luogo dalla mielografia. I risultati ottenuti indicano che il test di Lasègue è indicativo di compressione radicolare nel 90% dei casi, ma soltanto nei 2/3 dei casi la causa è ascrivibile all'ernia del disco. Un terzo dei pazienti è affetto da ernia del disco, ma con un SLR test negativo. *Pertanto l'SLR non ha un valore predittivo superiore ad altri test clinici.*

***Demircan M.N., Çolak A., Kutlay M., Kibıcı K., Topuz K. – Cramp finding: can it be used as a new diagnostic and prognostic factor in lumbar disc surgery<sup>37</sup>?***

L'obiettivo di questo studio prospettico è valutare se il "**Cramp Finding**", un nuovo tipo di test, può essere utilizzato come fattore diagnostico e prognostico nella chirurgia dell'ernia del disco. I risultati ottenuti dalla somministrazione del test sono stati confrontati con quelli della risonanza magnetica, del riscontro chirurgico (entrambi presi come golden standard) e dell'SLR, considerato il più importante test clinico per la valutazione della tensione a livello della radice nervosa lombare, causata da ernia del disco. E' stato valutato, in pratica, il grado di concordanza diagnostica tra risultati dell'SLR e quelli del Cramp Finding. Il Cramp Finding viene eseguito con il paziente in posizione prona; l'esaminatore chiede al paziente di effettuare una flessione di ginocchio massimale alla quale si oppone con una forza uguale e contraria. Il test è positivo se il paziente avverte un crampo fastidioso alla gamba o alla coscia. Il campione era costituito da 484 pazienti di età inferiore ai 45 anni. Il Cramp Finding è risultato positivo nel 72% dei pazienti operati, rispettivamente nel 70%, 52%, 34% e 8% dei pazienti in fase post-operatoria nel 1°, 3°, 12° e 24° mese. *Il Cramp Finding è assolutamente tollerato e sicuro per il paziente e sembra avere la stessa validità dell'SLR e, specialmente, dell'SLR controlaterale nella chirurgia dell'ernia del disco e nella valutazione dell'outcome. Il grado di concordanza diagnostica tra SLR e Cramp Finding, valutato con la costante di correlazione di Spearman, è risultato buono, con un valore pari a 0,284. Lo studio ha inoltre evidenziato valori di sensibilità e specificità del Cramp Finding rispettivamente pari al 72,3% ed al 100%. L'autore auspica ulteriori studi che valutino l'efficacia del Cramp Finding nella diagnosi di altre patologie, quali la spondilolisi, la stenosi, ecc.*

***Vroomen P.C., De Krom M.C., Wilmink J.T., Kester A.D., Knottnerus J.A. – Diagnostic value of history and physical examination in patients suspected of lumbosacral nerve root compression<sup>14</sup>.***

Lo studio di Vroomen si fissa come obiettivo di valutare le caratteristiche, i sintomi dei pazienti ed i risultati dell'esame fisico nella diagnosi clinica della compressione della radice nervosa lombosacrale che provoca la sciatica. Lo studio prende in considerazione 274 pazienti con dolore irradiato all'arto inferiore, ricoverati presso il reparto neurologico dell'Ospedale di Maastricht. I risultati evidenziano una correlazione tra la compressione radicolare e tre caratteristiche dei pazienti (sesso maschile; età compresa tra i 51 – 81 anni; tipo di lavoro che implichi camminare, stare in piedi e sollevare dei pesi), tre sintomi (tipico dolore dermatomero, sensazione di freddo nell'arto inferiore ed un incremento del dolore nello starnutare, tossire e durante gli sforzi) e quattro dati dell'esame fisico (paresi, assenza del riflesso patellare e tendineo, SLR test positivo ed un aumento della distanza dito-terra). *Tra gli elementi dell'esame fisico appena descritti, l'SLR test è quello che raggiunge il minor valore predittivo per una compressione radicolare, con valori di sensibilità e specificità inferiori a quanto riportato precedentemente in letteratura. L'SLR è in grado di indicare tensione o irritazione del nervo ma non necessariamente compressione della radice del nervo. Un incremento della distanza dito-terra (maggiore di 25 cm) e la paresi hanno invece un significativo valore predittivo per una compressione radicolare. Inoltre, lo studio evidenzia che la maggior parte delle informazioni diagnostiche fornite dall'esame fisico possono essere già reperite in fase di anamnesi clinica.*

***Albeck M.J. – A Critical Assessment of Clinical Diagnosis of Disc Herniation in Patients with Monoradicular Sciatica<sup>29</sup>.***

Albeck analizza il potere diagnostico dei test clinici, in particolare l'SLR ed il finger to floor distance test, per la valutazione dell'ernia del disco lombare nei pazienti con dolore monoradicolare a livello di L5-S1. Vengono inclusi nello studio 80 pazienti, ricoverati in un reparto di neurochirurgia. Il golden standard è rappresentato in questo caso dal riscontro chirurgico diretto. I risultati dei test diagnostici vengono interpretati attraverso la curva ROC. I valori di accuratezza diagnostica ottenuti dall'applicazione congiunta di entrambi i test sono i seguenti: sensibilità 0,95, specificità 0,88, valore predittivo positivo 0,96, valore predittivo negativo 0,85. Lo studio presenta diversi limiti: non differenzia i risultati derivanti dall'applicazione di ciascuno dei due test, ma li espone congiuntamente; vengono presentati soltanto i risultati dei pazienti dove è stata riscontrata chirurgicamente un'ernia del disco lombare (76%); non vengono presentati i risultati dei pazienti con risultato negativo ai test. Lo studio inoltre prende in considerazione un campione molto

specifico (pazienti ricoverati in un reparto di neurochirurgia), con conseguente alta probabilità che i test abbiano esito positivo. *Pertanto i risultati ottenuti, seppur rilevano livelli considerevoli di accuratezza diagnostica, non sono generalizzabili a livello di “cure primarie”, ossia di pazienti che ad esempio presentano per la prima volta nella loro vita segni e sintomi di sofferenza radicolare.* Nel campione analizzato, la localizzazione del dolore evocato con l'applicazione dei test consente di individuare correttamente il livello di ernia del disco nel 93% dei pazienti che presentano tale impairment. *Albeck evidenzia infine che il miglior indicatore per evidenziare l'ernia del disco lombare è il dolore irradiato e che l'integrazione tra l'anamnesi clinica ed esame fisico è la migliore strategia diagnostica per individuare pazienti affetti da radicolopatia.*

***Suri P., Rainville J., Katz J.N., Jouve C., Hartigan C., Limke J., Pena E., Li L., Swaim B., Hunter D.J. – The accuracy of the physical examination for the diagnosis of midlumbar and low lumbar nerve root impingement<sup>27</sup>.***

Suri analizza l'accuratezza dell'esame fisico per la diagnosi del conflitto radicolare a livello del rachide lombare medio e basso (L2-S1), utilizzando test individuali o combinazioni di test, confrontati con i risultati della risonanza magnetica. *I test clinici presi in considerazione sono principalmente lo straight leg raise (SLR) test normale ed incrociato (per la valutazione di L5-S1), il femoral stretch test (FST) normale ed incrociato (per la valutazione delle radici più alte), che presentano una sensibilità rispettivamente del 69% e del 50%.* Questi test, utilizzati singolarmente, possono fornire preziose informazioni cliniche che alterano sostanzialmente la probabilità post-test che sia presente il conflitto radicolare medio lombare nei pazienti che presentano dolore radicolare lombo-sacrale. Tuttavia, pur essendo caratterizzati da alta specificità, non raggiungono livelli significativi di sensibilità. Pertanto, l'utilizzo di test individuali nell'esame fisico potrebbe non rilevare molti casi di conflitto radicolare medio lombare. *L'utilizzo di test in combinazione mostra maggiori livelli di sensibilità. L'esame fisico può aiutare a distinguere quale radice nervosa è specificatamente coinvolta.* L'accuratezza diagnostica in questo studio è misurata principalmente con il rapporto di verosimiglianza o likelihood ratio (LR). Quest'ultimo è in grado di riassumere in un unico numero come la valutazione iniziale del rischio di malattia (probabilità pre-test) venga modificata da un risultato del test (probabilità post-test). Un LR positivo descrive come la probabilità di malattia venga modificata da un risultato positivo, mentre un LR negativo descrive come la probabilità di malattia venga modificata da un risultato negativo. Il rapporto di verosimiglianza, rispetto ad altri valori predittivi, ha il vantaggio di essere indipendente dalla prevalenza della malattia. I valori di likelihood ratio ottenuti sono stati i seguenti: a) SLR: LR+ 0,23

per L2, L3, L4 e 4,3 per L5 e S1; LR- 2,7 per L2, L3, L4 e 0,37 per L5 e S1; b) FST: LR+  $\infty$  per L2, L3, L4 e 0 per L5 e S1; LR- 0,50 per L2, L3, L4 e 2,0 per L5 e S1.

**Poiraudeau S., Foltz V., Drape J.L., Fermanian J., Lefevre-Colau M.M., Mayoux-Benhamou M.A. et al. – Value of the bell test and the hyperextension test for diagnosis in sciatica associated with disc herniation: comparison with Lasègue's sign and the crossed Lasègue's sign<sup>7</sup>.**

Lo studio di Poiraudeau ha analizzato il valore diagnostico di 4 test clinici nella diagnosi di sciatica associata ad ernia del disco: Bell test, hyperextension test, test di Lasègue normale ed incrociato. La risonanza magnetica e la tomografia computerizzata sono stati utilizzate come golden standard. I risultati evidenziano che il valore clinico del Bell test e dell'Hyperextension test è rilevante e quantomeno simile a quello del Lasègue test e del Lasègue test incrociato. *La combinazione tra il Lasègue test incrociato e l'hyperextension test ha eccellente specificità e un buon valore predittivo positivo per la diagnosi di sciatica associata all'ernia del disco. Pertanto il Bell Test e l'Hyperextension test andrebbero inseriti sistematicamente nel relativo esame fisico.*

**Kerr R.S.C., Cadoux-Hudson T.A., Adams C.B.T. – The value of accurate clinical assessment in the surgical management of the lumbar disc protrusion<sup>34</sup>.**

In questo studio 100 pazienti con protrusione discale lombare sono stati esaminati con l'SLR test normale e controlaterale e messi a confronto con il riscontro chirurgico (golden standard). I tre principali segni che indicavano una protrusione discale erano: SLR controlaterale positivo, atrofia del polpaccio e riduzione del riflesso achilleo (indicativa in particolare di una protrusione a livello L5-S1). Anche la debolezza nel movimento di dorsiflessione del piede e la scoliosi del rachide lombare erano segni importanti di protrusione discale. *La positività all'SLR test normale resta comunque il segno più frequente di protrusione discale; tuttavia è stata rilevata la sua positività anche in alcuni pazienti senza protrusione. Anche Kerr mette in evidenza come, ai fini dell'efficacia della valutazione di tali pazienti, sia necessario integrare l'anamnesi clinica con l'esame fisico.*

**Majlesi J., Togay H., Unalan H., Toprak S. – The Sensitivity and Specificity of the Slump and the Straight Leg Raising Testing in Patients With Lumbar Disc Herniation<sup>3</sup>.**

Majlesi nel suo studio effettua una comparazione tra la validità e l'accuratezza diagnostica dello Slump test e dell'SLR test. Il golden standard preso in considerazione è la risonanza magnetica. Lo Slump test è preferito all'SRL per diverse ragioni: 1) lo Slump è molto più sensibile perché aggiunge uno scivolamento craniale del nevrasse, mentre l'SRL esercita soltanto uno scivolamento caudale

delle radice del nervo; 2) lo Slump è più specifico perché la flessione ed estensione del collo aiutano a distinguere restrizioni di movimento del tessuto neurale da quelle di origine muscolo-scheletrica; 3) utilizzando l'SRL, è invece spesso difficile distinguere tra la tensione neurale e l'accorciamento dei muscoli della coscia e del gastrocnemio. Per ovviare a tale inconveniente è necessario chiedere al paziente un'accurata descrizione della localizzazione e della tipologia del suo dolore, non sempre facilmente definibile. Entrambi i test elicitano il dolore in presenza di ernia lombare, causata dalla trazione della radice coinvolta. L'SRL applica la propria tensione soprattutto a livello delle radici di L5-S1. Lo Slump invece può applicarla a tutte le radici lombari. Durante l'SLR, l'esaminatore esercita la tensione attraverso la flessione dell'anca e l'estensione del ginocchio. Durante lo Slump test questa è somministrata attraverso l'aggiunta della flessione delle colonna e della flessione d'anca. Lo Slump test è risultato più sensibile (0,84) dell'SLR (0,52) in pazienti affetti da ernia del disco. Tuttavia, l'SLR è lievemente più specifico dello Slump test, con valori rispettivamente pari a 0,89 ed a 0,83. Secondo l'autore, *lo Slump test è uno strumento dotato di alta sensibilità in pazienti con ernia del disco lombare che non producono compressione radicolare. Diversamente l'SLR, a causa della sua più alta specificità, può aiutare in particolar modo ad identificare pazienti con compressione radicolare, che richiedono un trattamento di tipo chirurgico.*

***Stankovic R., Johnell O., Maly P., Willner S. – Use of lumbar extension, slump test, physical and neurological examination in the evaluation of patients with suspected herniated nucleus pulposus. A prospective clinical study<sup>11</sup>.***

Stankovic nel suo studio valuta la validità del test dell'estensione lombare e dello Slump test nei pazienti con sospetta ernia lombare, confrontando i risultati dei test con quelli ottenuti con la risonanza magnetica e la Tomografia Computerizzata. Vengono esaminati 105 pazienti di un reparto di fisioterapia ortopedica (36 donne e 69 uomini, con età media di 42,7 anni, range 19-64 anni). *I valori congiunti di sensibilità e specificità dei due test clinici in rapporto alla valutazione radiologica sono risultati pari rispettivamente all'82,6% ed al 54,7%.* La concordanza diagnostica tra i risultati dei test clinici e quelli radiologici per tipo e livello di diagnosi di ernia del disco è risultata pari al 69%. Lo studio ha inoltre evidenziato che pazienti con ernia del disco hanno una riduzione significativa del ROM nella flessione del tronco partendo dalla stazione eretta ed un aumento del dolore irradiato lungo la gamba durante l'estensione del tronco. Sono necessari ulteriori studi per dimostrare se i due test esaminati possono essere considerati indicatori affidabili della presenza di ernia del disco lombare.

***Trainor K., Pinnington M.A. – Reliability and diagnostic validity of the slump knee bend neurodynamic test for upper/mid lumbar nerve root compression: a pilot study<sup>32</sup>.***

Trainor si pone l'obiettivo di determinare la validità diagnostica dello Slump test per le compressioni radicolari a livello lombare medio ed alto. Lo slump test è stato valutato da due esaminatori indipendenti e l'affidabilità inter-operatore è stata valutata con il coefficiente kappa. Il golden standard utilizzato è stato la risonanza magnetica e l'accuratezza diagnostica è stata misurata con i valori di sensibilità, specificità, valori predittivi positivo e negativo e con il likelihood ratio. Il campione era costituito da 16 pazienti con dolore radicolare, ricoverati in una clinica ortopedica specializzata in patologie del rachide (secondary care). I risultati hanno evidenziato un coefficiente kappa pari a 0,71 (livello di concordanza sostanziale), una sensibilità del 100%, una specificità dell'83%, un valore predittivo positivo del 67%, un valore predittivo negativo del 100%, un LR+ pari a 6,0 ed un LR- pari a 0. *Si può pertanto concludere che lo slump test è dotato da un buon livello di affidabilità inter-operatore e ha un potere diagnostico tale da poter essere considerato un test clinico molto utile per identificare pazienti con compressione radicolare a livello lombare medio.* Sono necessari ulteriori studi su un numero più consistente di pazienti per confermare tali risultati.

***Summers B., Mishra V., Jones J.M. – The flip test: a reappraisal<sup>33</sup>.***

L'obiettivo di questo studio è valutare la validità di costruito del **flip test**, come segno di tensione a livello della radice nervosa. Questo test è comunemente utilizzato in pazienti con sciatica, per confermare o meno una tensione a livello della radice nervosa, già evidenziata con un SLR test positivo. Il test è eseguito con il paziente seduto su un tavolo con ginocchia flesse, gambe penzoloni e con la schiena il più possibile estesa. L'esaminatore spinge con una mano la parte distale della coscia contro il tavolo e con l'altra afferra il calcagno e porta progressivamente la gamba in estensione. Il test è positivo se, oltre i 45° di estensione del ginocchio, si verifica un'inversione della lordosi lombare ed il paziente tende a cadere all'indietro con il tronco. Nonostante sia ampiamente noto, questo test non è ancora validato. Lo studio prende in considerazione 67 pazienti con sciatica, protrusione discale e compressione radicolare confermata dalla risonanza magnetica. Si sono evidenziati i seguenti risultati: il 33% dei pazienti non hanno sentito dolore durante l'esecuzione del test; il 39% ha sentito dolore solo in completa estensione di ginocchio ed il 28% dei pazienti ha opposto resistenza alla completa estensione del ginocchio a causa del dolore. Solo un terzo dei pazienti ha mostrato una tendenza alla caduta all'indietro del tronco. I risultati del flip test sono messi in correlazione con quelli dell'SLR. I pazienti con SLR

positivo sopra i 45° non hanno mostrato una reazione dolorosa nel flip test, mentre i pazienti con SLR positivo sotto i 45° hanno mostrato una risposta dolorosa nel flip test. Pertanto lo studio si conclude affermando *l'utilità del flip test nell'evidenziare tensione a livello della radice nervosa, ma solo nei pazienti con SLR positivo al di sotto dei 45°*. Il criterio di positività più affidabile non è la caduta all'indietro del tronco, ma l'elicitazione del dolore durante l'estensione del ginocchio.

***Christodoulides A.N. – Ipsilateral sciatica on femoral nerve stretch test is pathognomonic of an L4/5 disc protrusion<sup>36</sup>***

Lo studio di Christodoulides sostiene che il femoral nerve stretch test (di seguito FST) esercita una trazione sul nervo sciatico, sul nervo femorale, sul plesso e sulle radici lombosacrali. Il golden standard in questo caso è rappresentato dalla mielografia. Vengono studiati 200 pazienti con una sospetta protrusione discale a livello di L4-L5. In tali pazienti viene evocata una sciatica ipsilaterale con il FST, *che viene considerato un segno patognomico di protrusione discale laterale a livello di L4-L5*.

### **5.3: Valutazione qualità metodologica degli studi: Scala Quadas**

La qualità metodologica degli studi inclusi nella revisione è stata valutata con la Scala Quadas. Vengono di seguito riportati i risultati analitici.

1) Vroomen 2002 – Diagnostic value of history and physical examination in patients suspected of lumbosacral nerve root compression<sup>14</sup>. Final Score: 14/14  
 Test: SLR, Finger to floor distance test – Golden Standard (di seguito: G.S.): MRI

**Table 1**

The QUADAS tool

Item	Yes	No	Unclear
1. Was the spectrum of patients representative of the patients who will receive the test in practice?	(x)	( )	( )
2. Were selection criteria clearly described?	(x)	( )	( )
3. Is the reference standard likely to correctly classify the target condition?	(x)	( )	( )
4. Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?	(x)	( )	( )
5. Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?	(x)	( )	( )
6. Did patients receive the same reference standard regardless of the index test result?	(x)	( )	( )
7. Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?	(x)	( )	( )
8. Was the execution of the index test described in sufficient detail to permit replication of the test?	(x)	( )	( )
9. Was the execution of the reference standard described in sufficient detail to permit its replication?	(x)	( )	( )
10. Were the index test results interpreted without knowledge of the results of the reference standard?	(x)	( )	( )
11. Were the reference standard results interpreted without knowledge of the results of the index test?	(x)	( )	( )
12. Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?	(x)	( )	( )
13. Were uninterpretable/ intermediate test results reported?	(x)	( )	( )
14. Were withdrawals from the study explained?	(x)	( )	( )

2) Capra 2011 – Validity of the straight-leg raise test for patients with sciatic pain with or without lumbar pain using magnetic resonance imaging results as a reference standard<sup>31</sup>.  
Final Score: 13/14 – Test: SLR – G.S.: MRI

**Table 2**

The QUADAS tool

Item	Yes	No	Unclear
1. Was the spectrum of patients representative of the patients who will receive the test in practice?	(x)	( )	( )
2. Were selection criteria clearly described?	(x)	( )	( )
3. Is the reference standard likely to correctly classify the target condition?	(x)	( )	( )
4. Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?	( )	( )	(x)
5. Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?	(x)	( )	( )
6. Did patients receive the same reference standard regardless of the index test result?	(x)	( )	( )
7. Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?	(x)	( )	( )
8. Was the execution of the index test described in sufficient detail to permit replication of the test?	(x)	( )	( )
9. Was the execution of the reference standard described in sufficient detail to permit its replication?	(x)	( )	( )
10. Were the index test results interpreted without knowledge of the results of the reference standard?	(x)	( )	( )
11. Were the reference standard results interpreted without knowledge of the results of the index test?	(x)	( )	( )
12. Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?	(x)	( )	( )
13. Were uninterpretable/ intermediate test results reported?	(x)	( )	( )
14. Were withdrawals from the study explained?	(x)	( )	( )

**3) Suri 2011 – The accuracy of the physical examination for the diagnosis of midlumbar and low lumbar nerve root impingement<sup>27</sup>. Final Score: 13/14  
Test: SLR, CSLR, Femoral Stretch test (FST), Crossed FST – G.S.: MRI**

**Table 3**

The QUADAS tool

<b>Item</b>	<b>Yes</b>	<b>No</b>	<b>Unclear</b>
1. Was the spectrum of patients representative of the patients who will receive the test in practice?	( )	(x)	( )
2. Were selection criteria clearly described?	(x)	( )	( )
3. Is the reference standard likely to correctly classify the target condition?	(x)	( )	( )
4. Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?	(x)	( )	( )
5. Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?	(x)	( )	( )
6. Did patients receive the same reference standard regardless of the index test result?	(x)	( )	( )
7. Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?	(x)	( )	( )
8. Was the execution of the index test described in sufficient detail to permit replication of the test?	(x)	( )	( )
9. Was the execution of the reference standard described in sufficient detail to permit its replication?	(x)	( )	( )
10. Were the index test results interpreted without knowledge of the results of the reference standard?	(x)	( )	( )
11. Were the reference standard results interpreted without knowledge of the results of the index test?	(x)	( )	( )
12. Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?	(x)	( )	( )
13. Were uninterpretable/ intermediate test results reported?	(x)	( )	( )
14. Were withdrawals from the study explained?	(x)	( )	( )

4) Stankovic 1999 – Use of lumbar extension, slump test, physical and neurological examination in the evaluation of patients with suspected herniated nucleus pulposus. A prospective clinical study<sup>11</sup>. Final Score: 12/14 – Test: Slump Test – G.S.: MRI

**Table 4**

The QUADAS tool

Item	Yes	No	Unclear
1. Was the spectrum of patients representative of the patients who will receive the test in practice?	(x)	( )	( )
2. Were selection criteria clearly described?	(x)	( )	( )
3. Is the reference standard likely to correctly classify the target condition?	(x)	( )	( )
4. Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?	(x)	( )	( )
5. Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?	(x)	( )	( )
6. Did patients receive the same reference standard regardless of the index test result?	(x)	( )	( )
7. Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?	(x)	( )	( )
8. Was the execution of the index test described in sufficient detail to permit replication of the test?	(x)	( )	( )
9. Was the execution of the reference standard described in sufficient detail to permit its replication?	(x)	( )	( )
10. Were the index test results interpreted without knowledge of the results of the reference standard?	(x)	( )	( )
11. Were the reference standard results interpreted without knowledge of the results of the index test?	(x)	( )	( )
12. Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?	(x)	( )	( )
13. Were uninterpretable/ intermediate test results reported?	( )	( )	(x)
14. Were withdrawals from the study explained?	( )	( )	(x)

5) Rabin 2007 – The sensitivity of the seated straight-leg raise test compared with the supine straight-leg raise test in patients presenting with magnetic resonance imaging evidence of lumbar nerve root compression<sup>2</sup>. Final Score: 12/14 – Test: SLR – G.S.: MRI

**Table 5**

The QUADAS tool

Item	Yes	No	Unclear
1. Was the spectrum of patients representative of the patients who will receive the test in practice?	(x)	( )	( )
2. Were selection criteria clearly described?	(x)	( )	( )
3. Is the reference standard likely to correctly classify the target condition?	(x)	( )	( )
4. Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?	( )	( )	(x)
5. Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?	(x)	( )	( )
6. Did patients receive the same reference standard regardless of the index test result?	(x)	( )	( )
7. Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?	(x)	( )	( )
8. Was the execution of the index test described in sufficient detail to permit replication of the test?	(x)	( )	( )
9. Was the execution of the reference standard described in sufficient detail to permit its replication?	( )	(x)	( )
10. Were the index test results interpreted without knowledge of the results of the reference standard?	(x)	( )	( )
11. Were the reference standard results interpreted without knowledge of the results of the index test?	(x)	( )	( )
12. Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?	(x)	( )	( )
13. Were uninterpretable/ intermediate test results reported?	(x)	( )	( )
14. Were withdrawals from the study explained?	(x)	( )	( )

6) Majlesi 2008 – The sensitivity and specificity of the Slump and the Straight Leg Raising tests in patients with lumbar disc herniation<sup>3</sup>. Final Score: 11/14 – Test: SLR, Slump Test – G.S.: MRI

**Table 6**

The QUADAS tool

Item	Yes	No	Unclear
1. Was the spectrum of patients representative of the patients who will receive the test in practice?	( )	(x)	( )
2. Were selection criteria clearly described?	(x)	( )	( )
3. Is the reference standard likely to correctly classify the target condition?	(x)	( )	( )
4. Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?	( )	( )	(x)
5. Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?	(x)	( )	( )
6. Did patients receive the same reference standard regardless of the index test result?	(x)	( )	( )
7. Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?	(x)	( )	( )
8. Was the execution of the index test described in sufficient detail to permit replication of the test?	(x)	( )	( )
9. Was the execution of the reference standard described in sufficient detail to permit its replication?	(x)	( )	( )
10. Were the index test results interpreted without knowledge of the results of the reference standard?	(x)	( )	( )
11. Were the reference standard results interpreted without knowledge of the results of the index test?	(x)	( )	( )
12. Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?	(x)	( )	( )
13. Were uninterpretable/ intermediate test results reported?	( )	( )	(x)
14. Were withdrawals from the study explained?	(x)	( )	( )

7) Trainor 2011 – Reliability and diagnostic validity of the slump knee bend neurodynamic test for upper/mid lumbar nerve root compression: a pilot study<sup>32</sup>. Final Score: 11/14 – Test: Slump Test – G.S.: MRI

**Table 7**

The QUADAS tool

Item	Yes	No	Unclear
1. Was the spectrum of patients representative of the patients who will receive the test in practice?	( )	(x)	( )
2. Were selection criteria clearly described?	(x)	( )	( )
3. Is the reference standard likely to correctly classify the target condition?	(x)	( )	( )
4. Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?	( )	(x)	( )
5. Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?	(x)	( )	( )
6. Did patients receive the same reference standard regardless of the index test result?	(x)	( )	( )
7. Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?	(x)	( )	( )
8. Was the execution of the index test described in sufficient detail to permit replication of the test?	(x)	( )	( )
9. Was the execution of the reference standard described in sufficient detail to permit its replication?	( )	(x)	( )
10. Were the index test results interpreted without knowledge of the results of the reference standard?	(x)	( )	( )
11. Were the reference standard results interpreted without knowledge of the results of the index test?	(x)	( )	( )
12. Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?	(x)	( )	( )
13. Were uninterpretable/ intermediate test results reported?	(x)	( )	( )
14. Were withdrawals from the study explained?	(x)	( )	( )

8) Summers 2009 – The flip test: a reappraisal<sup>33</sup>. Final Score: 10/14 – Test: Flip Test – G.S.: MRI

**Table 8**

The QUADAS tool

Item	Yes	No	Unclear
1. Was the spectrum of patients representative of the patients who will receive the test in practice?	(x)	( )	( )
2. Were selection criteria clearly described?	(x)	( )	( )
3. Is the reference standard likely to correctly classify the target condition?	(x)	( )	( )
4. Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?	( )	( )	(x)
5. Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?	(x)	( )	( )
6. Did patients receive the same reference standard regardless of the index test result?	(x)	( )	( )
7. Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?	(x)	( )	( )
8. Was the execution of the index test described in sufficient detail to permit replication of the test?	(x)	( )	( )
9. Was the execution of the reference standard described in sufficient detail to permit its replication?	( )	( )	(x)
10. Were the index test results interpreted without knowledge of the results of the reference standard?	( )	( )	(x)
11. Were the reference standard results interpreted without knowledge of the results of the index test?	(x)	( )	( )
12. Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?	(x)	( )	( )
13. Were uninterpretable/ intermediate test results reported?	( )	( )	(x)
14. Were withdrawals from the study explained?	(x)	( )	( )

9) Poiraudeau 2001 – Value of the bell test and the hyperextension test for diagnosis in sciatica associated with disc herniation: comparison with Lasègue's sign and the crossed Lasègue's sign<sup>7</sup>. Final Score: 9/14 – Test: SLR, CSLR, BELL, HYPEREXT. – G.S.: MRI, CT

**Table 9**

The QUADAS tool

Item	Yes	No	Unclear
1. Was the spectrum of patients representative of the patients who will receive the test in practice?	( )	(x)	( )
2. Were selection criteria clearly described?	(x)	( )	( )
3. Is the reference standard likely to correctly classify the target condition?	(x)	( )	( )
4. Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?	( )	( )	(x)
5. Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?	(x)	( )	( )
6. Did patients receive the same reference standard regardless of the index test result?	( )	(x)	( )
7. Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?	(x)	( )	( )
8. Was the execution of the index test described in sufficient detail to permit replication of the test?	(x)	( )	( )
9. Was the execution of the reference standard described in sufficient detail to permit its replication?	( )	( )	(x)
10. Were the index test results interpreted without knowledge of the results of the reference standard?	(x)	( )	( )
11. Were the reference standard results interpreted without knowledge of the results of the index test?	(x)	( )	( )
12. Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?	(x)	( )	( )
13. Were uninterpretable/ intermediate test results reported?	(x)	( )	( )
14. Were withdrawals from the study explained?	( )	( )	(x)

**10) Albeck 1996 – A Critical Assessment of Clinical Diagnosis of Disc Herniation in Patients with Monoradicular Sciatica<sup>29</sup>. Final Score: 8/14 – Test: SLR, Finger to floor distance test – G.S.: Surgery**

**Table 10**

The QUADAS tool

<b>Item</b>	<b>Yes</b>	<b>No</b>	<b>Unclear</b>
1. Was the spectrum of patients representative of the patients who will receive the test in practice?	( )	( )	(x)
2. Were selection criteria clearly described?	(x)	( )	( )
3. Is the reference standard likely to correctly classify the target condition?	(x)	( )	( )
4. Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?	( )	( )	(x)
5. Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?	(x)	( )	( )
6. Did patients receive the same reference standard regardless of the index test result?	(x)	( )	( )
7. Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?	(x)	( )	( )
8. Was the execution of the index test described in sufficient detail to permit replication of the test?	(x)	( )	( )
9. Was the execution of the reference standard described in sufficient detail to permit its replication?	(x)	( )	( )
10. Were the index test results interpreted without knowledge of the results of the reference standard?	(x)	( )	( )
11. Were the reference standard results interpreted without knowledge of the results of the index test?	( )	(x)	( )
12. Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?	( )	( )	(x)
13. Were uninterpretable/ intermediate test results reported?	( )	( )	(x)
14. Were withdrawals from the study explained?	( )	( )	(x)

11) Xin 1987 – Significance of the straight-leg-raising test in the diagnosis and clinical evaluation of lower lumbar intervertebral-disc protrusion<sup>25</sup>. Final Score: 8/14 – Test: SLR – G.S.: Surgery

**Table 10**

The QUADAS tool

Item	Yes	No	Unclear
1. Was the spectrum of patients representative of the patients who will receive the test in practice?	(x)	( )	( )
2. Were selection criteria clearly described?	(x)	( )	( )
3. Is the reference standard likely to correctly classify the target condition?	( )	(x)	( )
4. Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?	( )	(x)	( )
5. Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?	( )	( )	(x)
6. Did patients receive the same reference standard regardless of the index test result?	(x)	( )	( )
7. Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?	(x)	( )	( )
8. Was the execution of the index test described in sufficient detail to permit replication of the test?	(x)	( )	( )
9. Was the execution of the reference standard described in sufficient detail to permit its replication?	( )	(x)	( )
10. Were the index test results interpreted without knowledge of the results of the reference standard?	( )	(x)	( )
11. Were the reference standard results interpreted without knowledge of the results of the index test?	(x)	( )	( )
12. Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?	(x)	( )	( )
13. Were uninterpretable/ intermediate test results reported?	( )	( )	(x)
14. Were withdrawals from the study explained?	(x)	( )	( )

12) Kerr 1988 – The value of accurate clinical assessment in the surgical management of the lumbar disc protrusion<sup>34</sup>. Final Score: 7/14 – Test: SLR, CSLR – G.S.: Surgery

**Table 10**

The QUADAS tool

Item	Yes	No	Unclear
1. Was the spectrum of patients representative of the patients who will receive the test in practice?	( )	(x)	( )
2. Were selection criteria clearly described?	( )	( )	(x)
3. Is the reference standard likely to correctly classify the target condition?	(x)	( )	( )
4. Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?	( )	( )	(x)
5. Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?	(x)	( )	( )
6. Did patients receive the same reference standard regardless of the index test result?	( )	(x)	( )
7. Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?	(x)	( )	( )
8. Was the execution of the index test described in sufficient detail to permit replication of the test?	( )	(x)	( )
9. Was the execution of the reference standard described in sufficient detail to permit its replication?	(x)	( )	( )
10. Were the index test results interpreted without knowledge of the results of the reference standard?	(x)	( )	( )
11. Were the reference standard results interpreted without knowledge of the results of the index test?	( )	( )	(x)
12. Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?	(x)	( )	( )
13. Were uninterpretable/ intermediate test results reported?	( )	( )	(x)
14. Were withdrawals from the study explained?	(x)	( )	( )

**13) Kosteljanetz 1984 – Predictive value of clinical and surgical findings in patients with lumbago-sciatica<sup>35</sup>. Final Score: 6/14 – Test: SLR – G.S.: Surgery, Myelography**

**Table 10**

The QUADAS tool

<b>Item</b>	<b>Yes</b>	<b>No</b>	<b>Unclear</b>
1. Was the spectrum of patients representative of the patients who will receive the test in practice?	( )	(x)	( )
2. Were selection criteria clearly described?	(x)	( )	( )
3. Is the reference standard likely to correctly classify the target condition?	( )	(x)	( )
4. Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?	( )	( )	(x)
5. Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?	(x)	( )	( )
6. Did patients receive the same reference standard regardless of the index test result?	(x)	( )	( )
7. Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?	(x)	( )	( )
8. Was the execution of the index test described in sufficient detail to permit replication of the test?	( )	(x)	( )
9. Was the execution of the reference standard described in sufficient detail to permit its replication?	( )	(x)	( )
10. Were the index test results interpreted without knowledge of the results of the reference standard?	( )	( )	(x)
11. Were the reference standard results interpreted without knowledge of the results of the index test?	( )	( )	(x)
12. Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?	(x)	( )	( )
13. Were uninterpretable/ intermediate test results reported?	( )	( )	(x)
14. Were withdrawals from the study explained?	(x)	( )	( )

**14) Christodoulides 1989 – Ipsilateral sciatica on femoral nerve stretch test is pathognomonic of an L4/5 disc protrusion<sup>36</sup>. Final Score: 5/14 – Test: Femoral Stretch Test – G.S.: Myelography**

**Table 10**

The QUADAS tool

<b>Item</b>	<b>Yes</b>	<b>No</b>	<b>Unclear</b>
1. Was the spectrum of patients representative of the patients who will receive the test in practice?	( )	(x)	( )
2. Were selection criteria clearly described?	( )	( )	(x)
3. Is the reference standard likely to correctly classify the target condition?	( )	(x)	( )
4. Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?	( )	( )	(x)
5. Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?	(x)	( )	( )
6. Did patients receive the same reference standard regardless of the index test result?	(x)	( )	( )
7. Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?	(x)	( )	( )
8. Was the execution of the index test described in sufficient detail to permit replication of the test?	(x)	( )	( )
9. Was the execution of the reference standard described in sufficient detail to permit its replication?	( )	(x)	( )
10. Were the index test results interpreted without knowledge of the results of the reference standard?	( )	( )	(x)
11. Were the reference standard results interpreted without knowledge of the results of the index test?	( )	( )	(x)
12. Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?	( )	( )	(x)
13. Were uninterpretable/ intermediate test results reported?	(x)	( )	( )
14. Were withdrawals from the study explained?	( )	( )	(x)

15) Demircan 2002 – Cramp finding: can it be used as a new diagnostic and prognostic factor in lumbar disc surgery<sup>37</sup>? Final Score: 5/14 – Test: SLR, CRAMP FIND. – G.S.: MRI, Surgery

**Table 10**

The QUADAS tool

Item	Yes	No	Unclear
1. Was the spectrum of patients representative of the patients who will receive the test in practice?	( )	( )	(x)
2. Were selection criteria clearly described?	(x)	( )	( )
3. Is the reference standard likely to correctly classify the target condition?	(x)	( )	( )
4. Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?	( )	( )	(x)
5. Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?	( )	(x)	( )
6. Did patients receive the same reference standard regardless of the index test result?	( )	(x)	( )
7. Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?	( )	( )	(x)
8. Was the execution of the index test described in sufficient detail to permit replication of the test?	( )	(x)	( )
9. Was the execution of the reference standard described in sufficient detail to permit its replication?	( )	(x)	( )
10. Were the index test results interpreted without knowledge of the results of the reference standard?	( )	( )	(x)
11. Were the reference standard results interpreted without knowledge of the results of the index test?	(x)	( )	( )
12. Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?	(x)	( )	( )
13. Were uninterpretable/ intermediate test results reported?	( )	( )	(x)
14. Were withdrawals from the study explained?	(x)	( )	( )

Per una valutazione schematica dei risultati ottenuti dall'applicazione della Scala Quadas agli studi inclusi nella revisione, è stata elaborata la seguente tabella di sintesi:

TEST	GOLDEN STD	AUTORE	TITOLO	VALUTAZIONE QUADAS
SLR + FINGER TO FLOOR TEST	MRI	Vroomen et al. 2002	Diagnostic value of history and physical examination in patients suspected of lumbosacral nerve root compression.	14/14
SLR	MRI	Capra F. et al. 2011	Validity of the straight-leg raise test for patients with sciatic pain with or without lumbar pain using magnetic resonance imaging results as a reference standard.	13/14
SLR, CSLR, FST, CFST	MRI	Suri P et al. 2011	The accuracy of the physical examination for the diagnosis of midlumbar and low lumbar nerve root impingement.	13/14
SLUMP TEST	MRI	Stankovic 1999	Use of lumbar extension, slump test, physical and neurological examination in the evaluation of patients with suspected herniated nucleus pulposus. A prospective clinical study.	12/14
SLR	MRI	Rabin A. et al. 2007	The sensitivity of the seated straight-leg raise test compared with the supine straight-leg raise test in patients presenting with magnetic resonance imaging evidence of lumbar nerve root compression.	12/14
SLR, SLUMP TEST	MRI	Majlesi J. et al. 2008	The sensitivity and specificity of the Slump and the straight leg raising tests in patients with lumbar disc herniation	11/14
SLUMP TEST	MRI	Trainor K et al. 2010	Reliability and diagnostic validity of the slump knee bend neurodynamic test for upper/mid lumbar nerve root compression: a pilot study.	11/14
FLIP TEST	MRI	Summers B et al. 2009	The flip test: a reappraisal.	10/14
LASÈGUE, CROSSED LASÈGUE, BELL TEST, HYPEREXTENSION TEST	MRI, CT	Poiraudeau et al. 2001	Value of the bell test and the hyperextension test for diagnosis in sciatica associated with disc herniation: comparison with Lasègue's sign and the crossed Lasègue's sign.	9/14
SLR + FINGER TO FLOOR TEST	SURGERY	Albeck 1996	A critical assessment of clinical Diagnosis of Disc Herniation in Patient with monoradicular sciatica	8/14
SLR	SURGERY	Xin SQ, Zhang QZ, Fan DH 1987	Significance of the straight-leg-raising test in the diagnosis and clinical evaluation of lower lumbar intervertebral-disc protrusion.	8/14
SLR, CSLR	SURGERY	Kerr 1988	The value of accurate clinical assessment in the surgical management of the lumbar disc protrusion	7/14
SLR	SURGERY, MIELOGRAPHY	Kosteljanetz et al. 1984	Predictive value of clinical and surgical findings in patients with lumbago-sciatica. A prospective study (Part I).	6/14
FEMORAL STRETCH TEST	MIELOGRAPHY	Christodoulidis A. 1989	Ipsilateral sciatica on femoral nerve stretch test is pathognomonic of an L4/5 disc protrusion.	5/14
SLR, CRAMP FINDING	SURGERY, MRI	Demircan et al. 2002	Cramp findings: can it be used as a new diagnostic and prognostic factor in lumbar disc surgery?	5/14

I test evidenziati nei riquadri in grassetto sono quelli con i punteggi più alti e che dispongono delle informazioni necessarie per l'elaborazione dell'F1 Score.

Lo studio di Vroomen è stato l'unico che ha ottenuto punteggio pieno alla valutazione con la Scala Quadas. Diversi studi hanno analizzato un campione molto specifico di soggetti, sottoposti ad esempio ad intervento chirurgico per la presenza di un ernia del disco lombare o ricoverati in reparti specializzati in patologie del rachide. Ovviamente la scelta del campione condiziona pesantemente i valori di accuratezza diagnostica dei test, ma rende i risultati poco generalizzabili, in particolar modo nel contesto delle cure primarie, ossia dei pazienti che presentano per la prima volta nella loro vita segni e sintomi compatibili con una sofferenza radicolare. Altro aspetto spesso carente negli studi è stata l'indicazione del tempo trascorso tra la somministrazione del test clinico e l'analisi strumentale (golden standard). Anche questo aspetto è rilevante ai fini dell'analisi comparata dei risultati dell'indagine clinica. In presenza di un ritardo consistente tra le due rilevazioni lo stato della patologia può modificarsi e falsare l'attendibilità del test clinico. La "consistenza" del ritardo è un concetto relativo al tipo di patologia in esame ed al suo stato di avanzamento. Ad esempio alcune settimane tra le due rilevazioni possono essere ininfluenti se la patologia è in fase cronica; viceversa in fase acuta anche pochi giorni possono modificare lo stato della patologia ed alterare i risultati della valutazione. Altro aspetto che talvolta è stato riscontrato carente negli studi in rapporto alla Scala Quadas è l'accurata descrizione delle modalità di esecuzione del test clinico e dello standard di riferimento, tale da permetterne la ripetizione. Alcune variazioni nella misura dell'accuratezza diagnostica possono scaturire da differenti modalità di esecuzione del test clinico e del golden standard. Inoltre, una descrizione poco chiara pregiudica la ripetibilità del test in contesti diversi da quello in esame. In più occasioni si è visto come differenze in termini di campione di riferimento, modalità di esecuzione del test, criteri di positività adottati abbiano modificato in maniera sostanziale i valori di sensibilità del test clinico.

## 5.4 Valutazione qualità psicometriche dei test: Tabella F1-Score

Test-Golden Std	Titolo	Valut. Quadas	Sens.	Spec.	PPV	NPV	LR +	LR-	F1 Score
<u>SLUMP</u> -MRI	The sensitivity and specificity of the Slump and the straight leg raising tests in patients with lumbar disc herniation	11/14	0,84	0,83	0,84	0,84	4,94	0,19	<u>0,84</u>
SLUMP-MRI	Reliability and diagnostic validity of the slump knee bend neurodynamic test for upper/mid lumbar nerve root compression: a pilot study.	11/14	1,00	0,83	0,67	1,00	5,88	0,00	0,80
LASÈGUE-MRI/CT	Value of the bell test and the hyperextension test for diagnosis in sciatica associated with disc herniation: comparison with Lasègue's sign and the crossed Lasègue's sign.	9/14	0,80	0,37	0,60	0,58	1,27	0,54	0,69
SLR-MRI	Diagnostic value of history and physical examination in patients suspected of lumbosacral nerve root compression.	14/14	0,64	0,57	0,65	0,56	1,49	0,63	0,64
CROSSED SLR-SURGERY	The value of accurate clinical assessment in the surgical management of the lumbar disc protrusion	7/14	0,43	0,94	0,96	0,37	7,17	0,61	0,59
FING. TO FLOOR-MRI	Diagnostic value of history and physical examination in patients suspected of lumbosacral nerve root compression.	14/14	0,45	0,74	0,68	0,52	1,73	0,74	0,54
BELL TEST-MRI/CT	Value of the bell test and the hyperextension test for diagnosis in sciatica associated with disc herniation: comparison with Lasègue's sign and the crossed Lasègue's sign.	9/14	0,46	0,63	0,59	0,49	1,24	0,86	0,52
HYPEREX TENSIO TEST-MRI/CT	Value of the bell test and the hyperextension test for diagnosis in sciatica associated with disc herniation: comparison with Lasègue's sign and the crossed Lasègue's sign.	9/14	0,44	0,67	0,61	0,49	1,33	0,84	0,51
CROSSED LASÈGUE-MRI/CT	Value of the bell test and the hyperextension test for diagnosis in sciatica associated with disc herniation: comparison with Lasègue's sign and the crossed Lasègue's sign.	9/14	0,29	0,83	0,68	0,49	1,71	0,86	0,41

Alla valutazione dell'accuratezza diagnostica dei test clinici effettuata con l'F1 score, lo Slump è risultato per due volte il test che ha ottenuto il punteggio più alto. E' stato scelto di sottoporre alla valutazione entrambi gli studi sullo Slump che presentavano il medesimo punteggio alla Scala Quadas (11/14). L'F1 Score è risultato pari rispettivamente a 0,84 e 0,80. Già lo studio di Majlesi<sup>3</sup>, comparando lo Slump Test con SLR, ne aveva evidenziato il maggior livello di sensibilità, di tensione esercitato a livello neurale e di applicabilità su tutte le radici nervose a livello lombare (mentre l'SLR ha una sensibilità maggiore a livello di L5-S1), la maggiore efficacia in termini di differenziazione strutturale tra problematiche di tensione a livello neurale piuttosto che di origine muscoloscheletrica. Segue subito dopo il test di Lasègue e l'SLR, con punteggi pari rispettivamente a 0,69 e 0,64. Da rilevare la marcata differenza tra il punteggio ottenuto dal crossed straight leg raise (CSLR) test, pari a 0,59 e quello ottenuto dal crossed Lasègue test, pari a 0,41, nonostante secondo la maggior parte degli autori siano sostanzialmente uguali. Ancora una volta tale differenza di accuratezza diagnostica è ascrivibile molto probabilmente alle metodologie di indagine adottate nei rispettivi studi, in termini di caratteristiche del campione di riferimento, modalità di esecuzione, criteri di positività, ecc. Come già evidenziato da Vroomen nel suo studio<sup>14</sup>, il finger to floor distance test è considerato un test molto affidabile ed ha ottenuto un punteggio pari a 0,54, molto simile a quello ottenuto dall'SLR. Andrebbe pertanto inserito come test standard nell'esame fisico per la valutazione delle problematiche di origine radicolare. Valori non molto distanti sono stati ottenuti anche dal Bell test e dall'hyperextension test, con punteggi pari rispettivamente a 0,52 e 0,51. Poiraudeau nel suo studio<sup>7</sup> evidenzia valori di accuratezza diagnostica simili al test di Lasègue ipsilaterale e controlaterale. Ancora più rilevanti risultano i valori di accuratezza diagnostica, in termini di specificità e valore predittivo positivo, ottenuti dalla combinazione dell'hyperextension test con il test di Lasègue controlaterale. Quest'ultima considerazione mette in rilievo che, seppur lo Slump risulti il test dotato del maggior livello di accuratezza diagnostica, la combinazione di più test clinici integrata con un'accurata anamnesi clinica rappresenta senz'altro la migliore "strategia diagnostica" per identificare una sindrome radicolare.

## 6. DISCUSSIONE

La letteratura revisionata evidenzia l'importanza di un'attenta valutazione dei pazienti con segni e sintomi di sofferenza radicolare, al fine di differenziarne l'origine neurale da quella muscoloscheletrica, individuare il/i livello/i coinvolto/i, la radice interessata ed impostare un piano di trattamento costruito ad hoc per lo specifico problema del paziente. Sono pertanto necessari test clinici dotati di un buon livello di accuratezza diagnostica che, congiuntamente alla fase anamnestica, si inseriscono all'interno del complesso processo di ragionamento clinico. Gli studi sull'accuratezza diagnostica dei test clinici per la valutazione delle sindromi radicolare nel low back pain non sempre forniscono al fisioterapista indicazioni esaustive e concordi, tali da uniformare le metodologie di applicazione ed interpretazione dei risultati ottenuti. Infatti, durante la revisione degli studi oggetto del presente lavoro, sono state individuate diverse criticità.

La maggior parte degli studi individuati sull'argomento analizza quasi esclusivamente le caratteristiche di affidabilità dei test clinici più comuni, principalmente dello Straight leg raise (SLR) test, nella versione ipsilaterale e controlaterale, del Lasègue test (se si vuole tener conto dei pochi elementi di differenziazione dall'SLR evidenziati da alcuni autori<sup>5,13</sup>). In secondo luogo vengono esaminati lo Slump test e la manovra di Wassermann (definita anche prone knee bending test o femoral stretch test), sempre nella versione omolaterale e controlaterale. Pochi studi hanno preso in considerazione test clinici meno conosciuti, ma non per questo meno validi. Vroomen ritiene addirittura che il finger to floor distance test sia un indicatore di compressione radicolare più affidabile dell'SLR, che indica tensione o irritazione della radice nervosa, ma non necessariamente compressione<sup>14</sup>. Poiraudau attribuisce al Bell test ed all'hyperextension test quantomeno le medesime qualità di accuratezza diagnostica dell'SLR normale ed incrociato. Pochi autori, infine, hanno introdotto elementi di novità per la valutazione delle problematiche di sofferenza radicolare a livello lombare. Summers nel suo studio descrive il flip test che, applicato congiuntamente all'SLR, può essere un efficace indicatore di tensione nervosa. Demircan analizza il "Cramp Finding", un test dotato di buoni livelli di sensibilità e specificità e di un alto grado di concordanza diagnostica con i risultati dell'SLR nell'individuazione di problematiche di sofferenza radicolare. Ci sono altri test, come ad esempio il Cecin's sign, il bowstring's sign o cram test, sui quali purtroppo non sono stati individuati studi che analizzassero la relativa accuratezza diagnostica.

Diversi studi sono molto datati e, per alcuni di essi, non è stato possibile il reperimento.

Alcuni studi non si occupano della descrizione del test in esame, altri lo fanno in maniera approssimativa. In altri studi lo stesso test viene descritto in maniera differente da diversi autori, i criteri di positività sono più o meno restrittivi e ciò compromette l'affidabilità interoperatore, il relativo grado di concordanza e la stessa accuratezza diagnostica. Questo è senz'altro uno dei motivi per cui si riscontrano valori di sensibilità, specificità, valore predittivo positivo e negativo e likelihood ratio sensibilmente differenti a parità di test analizzato.

La maggior parte degli studi prende in considerazione pazienti ricoverati in reparti specializzati in patologie della colonna o già sottoposti ad intervento chirurgico per ernia discale. Come già specificato precedentemente, ciò compromette i risultati dei test in termini di generalizzabilità ed accuratezza diagnostica. Sarebbe auspicabile condurre tali ricerche su campioni costituiti da pazienti che si sottopongono per la prima volta ad un'indagine diagnostica conseguentemente a segni e sintomi di sofferenza radicolare (cure primarie vs secondarie).

Diversi studi sono stati esclusi per la totale mancanza o inadeguatezza di uno standard di riferimento, in base al quale poter confrontare i valori di accuratezza diagnostica dei rispettivi test. Non c'è, inoltre, completo accordo su quale debba essere il golden standard per tali problematiche, anche se la maggior parte degli autori indica la risonanza magnetica come la più efficace per la diagnosi delle sindromi radicolari. Alcuni studi utilizzano, in alternativa o congiuntamente alla risonanza magnetica, la tomografia computerizzata, la mielografia ed il riscontro chirurgico diretto. Quest'ultima tecnica possiede un'accuratezza diagnostica elevatissima, tuttavia presenta lo svantaggio di poter essere applicata solo ad un campione ristretto ed altamente selettivo di pazienti che devono essere trattati con un intervento chirurgico e per i quali è molto più probabile che il test diagnostico abbia esito positivo. Inoltre, i pazienti con esito negativo ai test diagnostici preliminari all'intervento chirurgico ed alla diagnostica strumentale non potranno chiaramente essere valutati con il riscontro chirurgico<sup>5</sup>.

La maggior parte degli studi tratta di ernia/protrusione discale e non di sofferenza radicolare e, come sappiamo, l'ernia discale può essere solo una delle cause di compressione radicolare. Si è riscontrato come, in diverse occasioni, soggetti con ernia discale diagnosticata attraverso tecniche di imaging, risultino totalmente asintomatici. Sarebbe pertanto più corretto inserire le future ricerche in un quadro più ampio di sindrome radicolare, a prescindere dai singoli fattori eziologici.

Nelle tabelle che riassumono i risultati della valutazione con la Scala Quadas e con l'indice F1 Score, si nota che gli studi con valutazione più alta sono quelli che prendono in considerazione l'applicazione congiunta di più test clinici in batteria. Ciò non dev'essere considerato puramente casuale, ma rappresenta un elemento su cui riflettere. Infatti, sono numerosi i test descritti in letteratura per identificare una sindrome radicolare, ma non ne esiste nessuno che da solo sia in grado di individuarla correttamente. Di conseguenza, diversi autori hanno sottolineato la basilare necessità di fondare il processo di ragionamento clinico su due elementi:

- ✓ un'anamnesi clinica ben condotta è in grado di fornirci preziose indicazioni sulla patologia del paziente e di formulare le ipotesi diagnostiche maggiormente plausibili;
- ✓ un esame fisico ben strutturato, che possa avvalersi anche di batterie di test clinici che tengano conto delle informazioni ottenute nella fase precedente.

## 7. CONCLUSIONI

Nonostante le criticità evidenziate nella discussione, gli obiettivi del presente lavoro possono ritenersi raggiunti e numerosi sono gli spunti di riflessione.

Viene innanzitutto ridimensionata l'accuratezza diagnostica dello straight leg raise test per quanto resti, in ogni caso, un importante strumento nel processo di valutazione funzionale di una sindrome radicolare. La mancanza di uniformità nelle procedure di esecuzione e nei criteri di interpretazione dei risultati rendono altrettanto variabili i valori di accuratezza diagnostica. Inoltre, diversi studi evidenziano che l'SLR è un indicatore di tensione e conseguente irritazione della radice nervosa, piuttosto che di compressione radicolare. Quest'ultima viene evidenziata più efficacemente dallo Slump test e dal finger to floor distance test. Principalmente lo Slump viene considerato, in diversi studi, il test dotato della migliore accuratezza diagnostica per le problematiche a livello radicolare, tenendo anche conto dei punteggi ottenuti applicando l'F1 Score. Questo risultato costituisce una preziosa indicazione, di cui il fisioterapista può tener conto nella sua pratica clinica.

Non ci sono purtroppo ricerche su interessanti test clinici che vengono applicati per la valutazione della sindrome radicolare a livello lombare. Un esempio su tutti è costituito dal Kemp's test, di cui non esistono studi che analizzino i valori di accuratezza diagnostica. Ciò limita in modo consistente lo spettro di strumenti clinici basati su prove di efficacia a disposizione del fisioterapista, impoverendone il "potenziale diagnostico" e l'individuazione del percorso riabilitativo più efficace per il paziente. Anche il confronto tra i diversi test ne risulta impoverito, in quanto limitato ai pochi test finora oggetto di studio.

L'anamnesi è il punto di partenza fondamentale per la valutazione e la diagnosi della sindrome radicolare. L'utilizzo in maniera isolata di un test non sempre garantisce dei risultati attendibili. E' pertanto necessario utilizzare combinazioni di test clinici integrate con la fase anamnestica nell'ambito di un processo di ragionamento clinico più globale.

## 8. KEY POINTS

- ❖ L'accuratezza diagnostica di un test clinico si misura attraverso le sue qualità psicometriche, quali la sensibilità, la specificità, il valore predittivo positivo e negativo oppure avvalendosi di modelli quantitativi e statistici come ad esempio il rapporto di verosimiglianza o likelihood ratio, le curve di roc e l'F1 Score;
- ❖ Il **Golden Standard** (inteso come il migliore e più accettato metodo per identificare una malattia o la condizione di interesse) per valutare problematiche di intrappolamento della radice nervosa è rappresentato dalla **Risonanza Magnetica Nucleare**. In subordine vengono utilizzate anche la Tomografia Computerizzata, la Mielografia, l'esame discografico ed il riscontro chirurgico diretto;
- ❖ I test clinici che valutano una sofferenza radicolare possono agire attraverso differenti meccanismi: aumento della compressione radicolare, applicazione di una pressione meccanica diretta a livello della radice nervosa, aumento della tensione a livello delle strutture neurali. Questi ultimi vengono denominati "**test di neurotensione**"; tra questi troviamo lo slump test, lo straight leg raising test normale ed incrociato, il prone knee bending test;
- ❖ La maggior parte degli studi presenti in letteratura analizza l'accuratezza diagnostica dei test clinici in relazione a problematiche di ernia o protrusione discale e non in un quadro più ampio di sofferenza radicolare;
- ❖ Gli studi sull'accuratezza diagnostica dei test clinici per la valutazione delle sindromi radicolare nel low back pain acuto non sempre forniscono al fisioterapista indicazioni esaustive e concordi, tali da uniformare le metodologie di applicazione ed interpretazione dei risultati ottenuti;
- ❖ Lo **Slump test** è risultato il test dotato della migliore accuratezza diagnostica per le problematiche di sofferenza della radice nervosa in generale e di compressione radicolare in particolare. Lo **straight leg raise (SLR) test**, per quanto rimanga uno degli strumenti più importanti e maggiormente utilizzati nella valutazione delle radicolopatie, è soprattutto un indicatore di tensione e conseguente irritazione della radice nervosa, piuttosto che di compressione radicolare.

- ❖ Nella pratica clinica vengono utilizzati diversi test per la valutazione delle radicolopatie, per i quali non sussistono attualmente in letteratura prove di efficacia. Ciò limita in modo consistente lo spettro di strumenti clinici basati su prove di efficacia a disposizione del fisioterapista, impoverendone il *“potenziale diagnostico”* e l'individuazione del percorso riabilitativo più efficace per il paziente. Ne costituiscono un esempio il Kemp's test, il Cecin's sign ed il Bowstring's sign;
- ❖ L'anamnesi è il punto di partenza fondamentale per la valutazione e la diagnosi della sindrome radicolare. L'utilizzo in maniera isolata di un test non sempre garantisce dei risultati attendibili. E' pertanto necessario utilizzare combinazioni di test clinici integrate con la fase anamnestica nell'ambito di un processo di ragionamento clinico più globale.

## 9. BIBLIOGRAFIA

1. Govannoni S., Minozzi S., Negrini S. et al. – Diagnostic therapeutic flow-charts for low back pain patients: the Italian clinical guidelines, *Eura Medicophys* 2006; 42: 151-70.
2. Rabin A., Gerszten P.C., Karausky P., Bunker C.H., Potter D.M., Welch W.C. – The sensitivity of the Seated Straight-Leg Raise Test Compared With the Supine Straight-Leg Raise Test in Patients Presenting With Magnetic Resonance Imaging Evidence of Lumbar Nerve Compression. *Arch Phys Med Rehabil.* Vol. 88, July 2007.
3. Majlesi J., Togay H., Unalan H., Toprak S. – The Sensitivity and Specificity of the Slump and the Straight Leg Raising Testing in Patients With Lumbar Disc Herniation. *JCR* vol. 14, n. 2, Aprile 2008.
4. Atlas S.J., MD, MPH, Deyo R.A., MD, MPH. – Evaluating and managing acute low back pain in the primary care setting. *J. Gen. Intern. Med.* 2001; 16: 120-131.
5. Van der Windt D.A., Simons E., Riphagen II, Ammendolia C., Verhagen A.P., Laslett M., Devillé W., Deyo R.A., Bouter L.M., De Vet H.C., Aertgeerts B. – Physical examination for lumbar radiculopathy due to disc herniation in patients with lowback pain. *Cochrane Database Syst Rev.* 2010 Feb. 17;2.
6. Walsh J, Hall T. – Agreement and correlation between the straight leg raise and slump tests in subjects with leg pain. *J. Manipulative Physiol. Ther.* 2009 Mar-Apr; 32(3):184-92.
7. Poiraudéau S., Foltz V., Drape J.L., Fermanian J., Lefevre-Colau M.M., Mayoux-Benhamou M.A. et al. – Value of the bell test and the hyperextension test for diagnosis in sciatica associated with disc herniation: comparison with Lasègue's sign and the crossed Lasègue's sign. *Rheumatology (Oxford)* 2001;40:460–6.
8. Koes B.W., Van Tulder M.W., Peul W.C. – “Diagnosis and treatment of sciatica” *BMJ* 2007;334:1313-7.

9. Lis A.M., Black K.M., Korn H., Nordin M. – “Association between sitting and occupational LBP”, *Eur. Spine J* (2007) 16:283–298.
10. Rahman Shiri, Jaro Karppinen, Paivi Leino-Arjas, Svetlana Solovieva, Æ Helena Varonen, Æ Eija Kalso, Æ Olavi Ukkola, Æ Eira Viikari-Juntura – Cardiovascular and lifestyle risk factors in lumbar radicular pain or clinically defined sciatica: a systematic review, *Eur Spine J.* (2007) 16:2043–2054.
11. Stankovic R., Johnell O., Maly P., Willner S. – Use of lumbar extension, slump test, physical and neurological examination in the evaluation of patients with suspected herniated nucleus pulposus. A prospective clinical study. *Man. Ther.* 1999;4:25–32.
12. Lezioni del Master in Riabilitazione dei Disordini Muscolo Scheletrici dell’Università degli studi di Genova, Facoltà di Medicina e Chirurgia, anno accademico 2011/2012.
13. LG PNLG 2005 – “Appropriatezza della diagnosi e del trattamento chirurgico dell’ernia del disco lombare sintomatica”.
14. Vroomen P.C., De Krom M.C., Wilmink J.T., Kester A.D., Knottnerus J.A. – Diagnostic value of history and physical examination in patients suspected of lumbosacral nerve root compression. *J. Neurol. Neurosurg. Psychiatry* 2002; 72:630–4.
15. Pedrabissi L. et al., *Test Psicologici, Esercitazioni Guidate: il concetto di attendibilità*, 1997, 1-2.
16. Cleland J., *L’esame clinico ortopedico, un approccio EBM*, 2006 masson ed., cap. 1, 3-15.
17. Urbano P., *Linee guida per la valutazione dei test diagnostici*, 2009 Jan.; 2-4.
18. Laslett M. – “Evidence-based diagnosis and treatment of the painful sacroiliac joint” *The Journal of Manual & Manipulative Therapy* 2008;16(3):142-152

19. Bottarelli E., Parodi S., Un approccio per la valutazione della validità dei test diagnostici: le curve R.O.C. (Receiver Operating Characteristic), Ann. Fac. Medic. Vet. di Parma (Vol. XXIII, 2003), 49-68.
20. Ozlem Uzuner, Imre Solti, Eithon Cadag, Extracting medication information from clinical text, 2010 June, 1-2.
21. Modic T. et al. – Lumbar Herniated Disk Disease and Canal Stenosis: Prospective Evaluation by Surface Coil MR, CT and Myelography, AJR, October 1986, 147:757-765.
22. Gibson M.J. et al. - Magnetic Resonance Imaging and discography in the diagnosis of disc degeneration: a comparative study of 50 discs, British Editorial Society of Bone and Joint Surgery, May 1986, Vol. 68 B, n° 3, 1-5.
23. Jarvik J.G., Deyo R.A. – Diagnostic evaluation of low back pain with emphasis on imaging. Ann Intern Med 2002; 137: 586-97.
24. Linee guida per la diagnostica per immagini. Documento approvato dalla Conferenza Stato Regioni il 28 Ottobre 2004, Repertorio Atti n. 2113.
25. Xin S.Q., Zhang Q.Z., Fan D.H. – Significance of the straight-leg-raising test in the diagnosis and clinical evaluation of lower lumbar intervertebral-disc protrusion. J. Bone Joint Surg. Am. 1987 Apr; 69(4):517-22.
26. Atlante delle tecniche di terapia manuale del Master in Riabilitazione dei Disordini Muscolo Scheletrici dell'Università degli studi di Genova, Facoltà di Medicina e Chirurgia, anno accademico 2011/2012.
27. Suri P., Rainville J., Katz J.N., Jouve C., Hartigan C., Limke J., Pena E., Li L., Swaim B., Hunter D.J. – The accuracy of the physical examination for the diagnosis of midlumbar and low lumbar nerve root impingement. Spine (Phila Pa 1976). 2011 Jan. 1;36(1):63-73.

28. Perret C., Poiraudreau S., Fermanian J., Colau M.M., Benhamou M.A., Revel M. – Validity, reliability, and responsiveness of the fingertipto-floor test., *Arch. Phys Med. Rehabil.* 2001;82:1566-70.
29. Albeck M.J. – A Critical Assessment of Clinical Diagnosis of Disc Herniation in Patients with Monoradicular Sciatica, *Acta Neurochir. (Wien)* (1996) 138:4044.
30. Whiting P., Rutjes A.W.S., Reitsma J.B., Bossuyt P.M.M. , Kleijnen J. – The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Medical Research Methodology* 2003, 3: 25 doi:10.1186/1471-2288-3-25.
31. Capra F., Vanti C., Donati R., Tombetti S., O'Reilly C., Pillastrini P. – Validity of the straight leg raise test for patients with sciatic pain with or without lumbar pain using magnetic resonance imaging results as a reference standard. *J. Manipulative Physiol. Ther.* 2011 May;34(4):231-8.
32. Trainor K., Pinnington M.A. – Reliability and diagnostic validity of the slump knee bend neurodynamic test for upper/mid lumbar nerve root compression: a pilot study. *Physiotherapy* – 2011 Mar.; 97(1):59-64.
33. Summers B., Mishra V., Jones J.M. – The flip test: a reappraisal – *Spine (Phila Pa 1976)*. 2009 Jul. 1;34(15):1585-9.
34. Kerr R.S.C., Cadoux-Hudson T.A., Adams C.B.T. – The value of accurate clinical assessment in the surgical management of the lumbar disc protrusion. *Journal of Neurology, Neurosurgery, and Psychiatry* 1988; 51:169-173.
35. Kosteljanetz M., Espersen J.O., Halaburt H., Miletic T. – Predictive value of clinical and surgical findings in patients with lumbago-sciatica. A prospective study (Part I). *Acta Neurochir. (Wien)*. 1984; 73(1-2):67-76.

36. Christodoulides A.N. – Ipsilateral sciatica on femoral nerve stretch test is pathognomonic of an L4/5 disc protrusion. *J. Bone Joint Surg. Br.* 1989 Jan; 71(1):88-9.
37. Demircan M.N., Çolak A., Kutlay M., Kıbıcı K., Topuz K. – Cramp finding: can it be used as a new diagnostic and prognostic factor in lumbar disc surgery? – *Eur. Spine J.* (2002) 11:47–51.