

PROJECTE DE RECERCA

Fiabilitat inter/intraexaminador dels tests osteopàtics

Revisió bibliogràfica

Autora: Neus Olivas García

nolivas@a-ver.es

Escola D'Osteopatia de Barcelona

Sant Just Desvern

27/01/2012

Supervisor Personal del Projecte: David Sánchez Rienda

Agraïments

Cal no abandonar mai ni la tasca ni l'esperança.

Pompeu Fabra

Vull agrair la col·laboració de:

El meu tutor, David Sánchez, per la llibertat i facilitat de tracte durant la planificació, i els consells i suport en la recta final de la confecció d'aquest projecte.

Els professors de Metodologia de la Recerca pels coneixements i informació que ens han facilitat per poder realitzar aquest treball. Especialment a Montse Serra pels consells i resolució de dubtes bibliogràfics.

L'Alba, la Montse, l'Ainhoa, la Vanessa i l'Olga per llegir, corregir i donar la seva sincera opinió.

El Carles pel seu ajut tècnic en la realització del gràfics estadístics, per la paciència i el suport personal i humà.

Resum

Introducció: En l'exploració física es realitzen tota una sèrie de tests. Els tests han de ser vàlids i fiables. La fiabilitat és la mesura de confiança del correcte funcionament d'un sistema i pot ser tant intra com interexaminador.

Material i mètode: Es va realitzar una recerca d'articles en bases de dades electròniques i recerca manual des de setembre del 2010 fins desembre del 2011. Es va avaluar la qualitat dels articles amb el sistema QUADAS i es van analitzar els resultats estadístics dels valors k.

Resultats: Es localitzen 183 articles dels que s'estudien 25 segons els criteris d'inclusió. Qualitativament van donar un resultat mig de 7,96 sobre 14. Del total d'articles 17 presenten dades de fiabilitat interexaminador i 7 intraexaminador.

Discussió: La qualitat dels articles és acceptable segons el sistema QUADAS. Els tests osteopàtics intenten determinar la presència d'una disfunció somàtica segons el criteri STAR. La fiabilitat intraexaminador de la pelvis en conjunt no arriba a k 0,4. La fiabilitat interexaminador lumbar té un valor k en conjunt de 0,31 i a la pelvis 0,21.

Conclusió: No hi ha cap test osteopàtic dels estudiats amb valors acceptables de fiabilitat. La fiabilitat intraexaminador és més alta que la interexaminador però totes amb nivells de valor k baixos.

Paraules Clau: Test osteopàtic, fiabilitat interexaminador, fiabilitat intraexaminador, validesa.

Abstract

Introduction: Some tests are conducted in the physical exploration. Those tests need to be valid and reliable. Reliability, which can be both interexaminer and intraexaminer, is a measure of trust used to check the working order of the tests.

Material and method: A research of articles was conducted in different electronic and manual databases from September 2010 to December 2011. The quality of the articles was evaluated with QUADAS' system and the statistics results of the k values were analyzed.

Results: From the 183 articles found, 25 were studied according to the established criteria. In quality terms, the average result was 7,96 out of 14. 17 of the 25 articles studied present interexaminer reliability data. Moreover, 7, out of the last 17 articles mentioned, present intraexaminer reliability data.

Discussion: The quality of the articles is acceptable, according to QUADAS' system criteria. Osteopathic tests try to determine the presence of a somatic dysfunction, according to STAR criteria. The overall pelvis intraexaminer reliability is below k 0,4. The lumbar interexaminer reliability has an overall k result of 0,31; while pelvis interexaminer reliability is 0,21.

Conclusion: Taking into account the studied tests, there is not any osteopathic tests with acceptable values of reliability. Intraexaminer reliability is higher than interexaminer reliability. Yet, both of them present low levels of k values.

Key words: Osteopathic test, interexaminer reliability, intraexaminer reliability, validity.

Índex

Continguts:

Certificació	I
Pàgina del títol	II
Agraïments	III
Resum	IV
Abstract	V
Llistat de figures i taules	VII
Llistat de gràfics	VIII
Llistat d'abreviatures	IX
Introducció	1
Material i mètode	
Selecció d'articles	5
Avaluació de la qualitat	7
Extracció de dades	8
Anàlisi estadística	8
Planificació de la recerca	10
Resultats	11
Discussió	17
Conclusions	28
Bibliografia	29
Annexes	40

Llistat de figures i taules

Figures

Figura 1: Diagrama de flux (Flow chart) d'inclusió dels articles 6a

Taules

Taula 1: Adaptació del sistema QUADAS d'avaluació de la qualitat de la precisió dels articles. 39

Taula 2: Mesures d'acord per dades categòriques segons Landis & Koch. 9a

Taula 3: Descriptors dels coeficients de correlació segons Hopkins. 9b

Taula 4: Cronograma de planificació de les tasques del projecte. 10a

Taula 5: Quadre de valoració qualitativa segons criteri QUADAS 41

Taula 6: Característiques dels estudis inclosos distribuïts per regions i en ordre alfabètic del primer autor. 42

Llistat de gràfics

Gràfic 1: Fiabilitat interexaminador lumbar	14a
Gràfic 2: Fiabilitat interexaminador pelvis	15a
Gràfic 3: Fiabilitat intraexaminador pelvis	16a

Llistat d'abreviatures

♀: Gènere femení

♂: Gènere masculí

AIL: Angle Infero-Lateral del Sacre

AILa/p: Angle Infero-Lateral del Sacre anterior/posterior

AILi: Angle Infero-Lateral del Sacre inferior

AILp: Angle Infero-Lateral del Sacre posterior

AILs/i: Angle Infero-Lateral del Sacre superior/inferior

AOA: American Osteopathic Association

Asim: Asimetria

ATM: Articulació Temporo Mandibular

Bp: Bipedestació

C: Vertebra Cervical

CCI: Coeficient de Correlació Intraclasse

D: Dreta

DP: Decúbit Pro

DS: Disfunció Somàtica

E: Esquerre

EIAS: Espina Ilíaca Antero-Superior

EIPI: Espina Ilíaca Postero-Inferior

EIPS: Espina Ilíaca Postero-Superior

EOB: Escola d'Osteopatia de Barcelona

EPS: Espina Postero Superior

K: Coeficient Kappa

L: Vertebra Lumbar

LF: Latero Flexió

MESH: Medical Subjects Headings

Met: Mètode

MM: Mal·lèol Medial

MRP: Moviment Respiratori Primari

PC: Personal Computer, Ordinador

PFP: Proba de Flexió Posterior

Post grup: Post entrenament grup

Post in: Post entrenament individual

Pre grup: Pre entrenament grup

Pre ind: Pre entrenament individual

Press: Pressió

QUADAS: Quality Assessment of Studies of Diagnostic Accuracy Included in Systematic Review

RL: Rebot Lumbar

Rot: Rotació

RX: Radiografia

Sd: Sedestació

Sens: Sensibilitat

SI: Sacre Ilíaca

SS: Solc Sacre

STAR: Sensibilitat alterada, anormalitat en la Textura dels Teixits, Asimetria de la forma, Restricció dels moviments.

STARDS: Standards for Reporting of Diagnostic Accuracy

T: Vèrtebra Toràtica

TA: Test Actius

TART: "Tissue Texture, Abnormal Quality of Resistance, Restriction of Vertebral Motion and Tenderness".

TC: Tomografia Computeritzada

TDPARCI: "Test de Deslizamiento Postero-Anterior del Raquis Cervical Inferior"

Teix: Teixits, alteració dels teixits

TFA: Test de Flexió Anterior

TMFET: Test de Mobilitat Flexo-Extensió del Turmell

TPS: Tibio-Peroneo Superior

UOC: Universitat Oberta de Catalunya

UV: Ultraviolat

Var: Variables

VPN: Valor Predictiu Negatiu

VPP: Valor Predictiu Positiu

Introducció

A la pràctica clínica en osteopatia abans de l'exploració física d'un pacient s'inicia un procés, on mitjançant l'anamnesi, es recullen dades subjectives que determinaran aquesta exploració per arribar mitjançant diversos tests a una hipòtesi de treball. A l'exploració es realitza una observació, palpació i tests tan actius com passius. Aquests tests, com en altres disciplines, mai confirmen o exclouen de forma absoluta la presència d'una hipòtesi, sinó que determinen la probabilitat que existeixi una alteració. Un cop recollida tota la informació és quan s'analitzen les dades per planificar la nostra intervenció i per tant, el tipus de tractament.

Això ens porta a reclamar cert grau de certesa sobre la informació que ens aporten els tests per què sigui vàlida i fiable, o millor dit com diu Cleland¹ "reduir el grau d'incertesa". Per tal de realitzar un diagnòstic precís i un tractament eficient i eficaç els tests han de ser el més vàlids i fiables possibles.

Com continua dient Cleland¹ "un test es considera fiable si proporciona informació precisa, concreta i reproduïble, i té validesa diagnòstica si té la capacitat de discriminar entre pacients amb i sense una alteració específica". La fiabilitat és un grau de consistència amb què es mesura un atribut en particular. I hi ha dos tipus de fiabilitat: la fiabilitat intraexaminador i la fiabilitat interexaminador:

- La fiabilitat intraexaminador és la mesura de la capacitat d'un mateix observador per realitzar mesures idèntiques en execucions separades del mateix test.

- La fiabilitat interexaminador és la mesura de la capacitat de dos o més observadors per obtenir resultats idèntics amb el mateix test.

Existeixen molts articles publicats sobre la validesa i la fiabilitat tant intra com interexaminador amb diferent tipus de resultats i variables. Tots aquests paràmetres de fiabilitat i validesa es presenten de manera numèrica i són el resultat de càlculs estadístics^{2,3} concrets que donen diferents variables que s'interpreten de manera diferent⁴ per donar sentit a l'objecte de la recerca en els diversos estudis.

És important que els qui realitzen els estudis escullin correctament el tipus de càlculs i variables estadístiques adequades segons el tipus de dades recollides que es volen analitzar³⁻⁵.

En resum podem dir que es poden recollir dades categòriques (nominals o ordinals) i dades no categòriques (d'interval o raó). Per les categòriques s'aconsella fer servir el *percentatge de concordança* per determinar el percentatge de vegades que els resultats obtinguts pels terapeutes són idèntics en un test; o bé el *coeficient kappa* (k) que és una mesura de proporció de l'acord potencial després d'eliminar l'atzar. Si les dades són d'interval o raó es solen utilitzar coeficients de correlació per determinar la fiabilitat. Com exemples tenim el *coeficient de correlació de Pearson* (r) i el *coeficient de correlació intraclasse* (CCI).

Hi ha autors com Lucas i Bogduk³ que determinen directament quins valors s'han de fer servir i aposten per utilitzar el CCI per dades contínues (com l'angle de mobilitat d'una articulació o la pressió arterial) i kappa per tests que generen dades categòriques (com positiu/negatiu o mig/moderat/sever). De totes formes, continuen argumentant que tot i que molts tests recullen dades contínues com la diferència en mil·límetres d'una dismetria, s'acaben traduint a dades categòriques com absent/present, amb lo que l'ús de kappa és el recomanat; fins i tot, el coeficient kappa ens és útil encara que es pugui veure "alterat" en casos que el signe clínic o la condició que es busca sigui massa gran o massa petita en la mostra d'estudi. Llavors recomanen l'ús de kappa ajustat i/o ponderat i adjuntar la raó de probabilitat positiva i la raó de probabilitat negativa per donar un context estadístic.

Des de l'osteopatia i d'altres disciplines manuals existeixen estudis experimentals per determinar el grau de fiabilitat de diversos tests que valoren si procediments habituals que es fan servir en la pràctica clínica diària són vàlids o no⁶⁻¹⁵. Cada vegada hi ha estudis més ben definits i que es porten a terme amb una metodologia molt acurada¹⁶⁻²¹. De totes formes la necessitat de dotar de fiabilitat alguns procediments porta a estudis que no estan dissenyats amb tota la cura que caldria per què el seu objectiu d'estudi i les seves conclusions siguin preses en consideració.

Des de diverses disciplines de la teràpia manual hi ha autors amb molta publicació sobre la fiabilitat com és el cas de van de Wurff²²⁻²⁴ Fryer²⁵⁻²⁸ o Laslett^{29,30}. En el cas de van de Wurff o de Laslett es realitzen sobretot estudis de provocació de dolor. Fryer busca aspectes més qualitius en la recerca des del que podríem dir que és un procediment més osteopàtic.

La diferència entre uns estudis i altres recau sobretot que els test des de la fisioteràpia i/o la teràpia física van a buscar el signe o símptoma directament provocant-lo, i en els tests més osteopàtics, sobretot els basats en el model teòric de Fred Mitchell³¹, busquen dismetries i/o anormalitats en el teixit que corresponen a la definició de criteri diagnòstic de disfunció somàtica amb l'acrònim STAR³²⁻³⁴ i que serien:

- Sensibilitat alterada
- Canvis en la textura dels teixits
- Asimetria posicional
- Restricció de la mobilitat

Existeixen estudis en la literatura científica per validar tests posturals³⁵⁻³⁸, tests ortopèdics^{39,40}, test palpatoris⁷... Per validar procediments⁴¹⁻⁴⁶ o fins i tot per validar la fiabilitat de la palpació de la respiració primària cranial⁴⁷⁻⁴⁹ segons el concepte d'osteopatia cranial descrit per Sutherland.

També existeixen molts estudis sobre la fiabilitat de determinats test^{11,13,50} i procediments, i moltes revisions literàries^{41,51-56}, però poques que separin exclusivament aquells tests realitzats des de i per l'osteopatia.

Per altra banda hi ha molts procediments i tests realitzats des d'altres disciplines que ens són útils i que fem servir diàriament, però aquestes revisen i valoren recerques científiques tant pròpies com alienes^{51,57-60}.

En l'àmbit exclusivament osteopàtic poques revisions s'han trobat on hi hagi una recerca completa i acurada no només dels estudis, sinó també de la qualitat d'aquest. Al 2002 Vaughan⁵⁴ va realitzar una revisió sobre les disfuncions cervicals i al 2010 Stovall⁶¹ va publicar una altra sobre asimetries de referències òssies a la regió lumbopèlvica. No s'han localitzat en publicacions d'osteopatia o en altres, revisions només d'estudis realitzats per osteòpates.

El que si existeix però des de l'osteopatia són molts articles "convidant" a fer servir els tests segons la seva utilitat clínica⁶², reclamant fiabilitat diagnòstica³ o explicant quin tipus de variables fer servir i com calcular-les per definir la utilitat dels tests⁵.

Potser és el moment de buscar en els fons bibliogràfics per revisar els resultats de diferents articles científics publicats des de l'osteopatia ordenar-los, avaluar la seva qualitat i sobretot definir quina fiabilitat tenen els tests i/o procediments que allà s'estudien.

Per això es realitza aquesta Revisió Bibliogràfica que té el modest objectiu de delimitar la qualitat dels estudis segons criteris publicats, definir que són els "tests osteopàtics" i el seu nivell fiabilitat inter-examinador i intra-examinador.

Selecció d'articles

Amb la finalitat de localitzar el màxim d'articles rellevants, es va fer una recerca de forma individual durant els mesos de setembre 2010 fins a desembre 2011 on es van incloure tots aquells articles que complissin les següents condicions:

- Els resultats havien d'estar publicats com a text complet abans de desembre del 2011 (els resums –*abstracts*- no es van incloure)
- Els estudis havien de tractar de *validesa*, *fiabilitat inter-examinador*, *fiabilitat intra-examinador* o *fiabilitat inter e intra-examinador*.
- Els estudis havien de ser publicats en anglès, castellà o català. Textos en alemany, francès o altres llengües van ser exclosos.

Per realitzar la recerca es van fer servir buscadors i bases de dades electròniques com Medline (via Pubmed), Google Scholar, Elsevier, Science Direct, Scirus, BioMed Central, Osteopathic Research, Trip Database (en castellà), Osteomed Dr. Osteopathic Medicine Digital Repository, CINAHL (des de la biblioteca virtual de la UOC), SUMSearch.

Per realitzar aquesta recerca s'ha fet servir vocabulari lliure i Tesauro Mesh (Medical Subjects Headings) que inclou: "Osteopathic Medicine" [Mesh], "Physical Examination" [Mesh], "Diagnostic Test" [Mesh], "Reproducibility of Results" [Mesh], "Osteopatia", "Test osteopático", "Tests osteopáticos", "Validez", "Fiabilidad", "Interexaminador", "Inter examinador", "Intraexaminador", "Intra examinador", "Palpación", "Osteopathic test", "Reliability".

De forma manual es va fer una recerca en llibres i revistes a la biblioteca de l'EOB, així com de les referències localitzades als diferents articles i

revisions bibliogràfiques i de les revistes més significatives en el tema com són: International Journal of Osteopathic Medicine, Journal of the American Osteopathic Association, Chiropractic and Osteopathy, Journal of Bodywork and Movement Therapies, Journal of Manipulative and Physiological Therapeutics, Journal of Osteopathic Medicine, Physical Therapy, Manual Therapy, Australian Journal of Physiotherapy, Journal of Strength and Conditioning Research.

Es van localitzar un total de 183 articles potencialment útils (tant en recerca manual com electrònica) dels quals es van excloure 100 per no tenir a veure amb el tema a estudiar. Dels 83 restants, 22 es van excloure per ser revisions bibliogràfiques , conferències o masterclass⁶² i 36 per no complir els criteris d'inclusió o no ser test osteopàtics o estudis no realitzat per osteòpates. Aquí caldria esmentar que 11 d'aquests^{6,11,13,57,63-69} són realitzats per fisioterapeutes amb formació de postgrau en teràpies manipulatives que per analogia es podrien haver inclòs dins l'estudi. Ja que es va decidir estudiar només articles publicats per osteòpates aquests finalment es van quedar fora de la selecció. Al final es van estudiar un total de 25 articles que són els que componen l'estudi. (Figura 1).

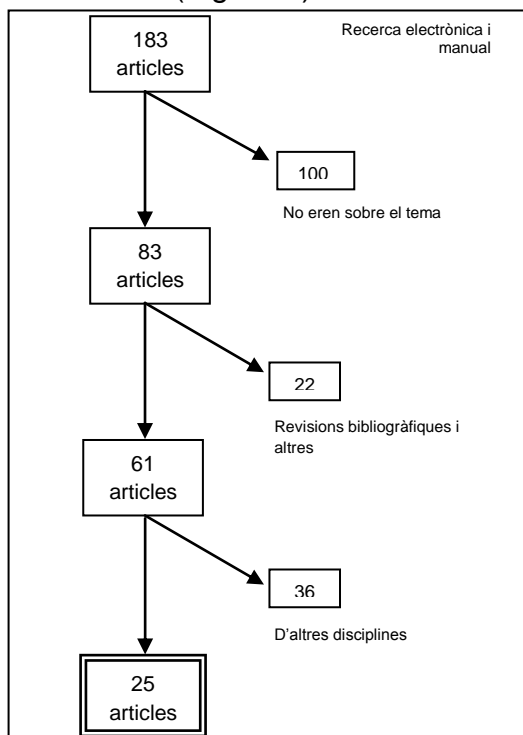


Figura 1: Diagrama de flux (Flow chart) d'inclusió dels articles

Aquests 25 es distribueixen de la següent forma:

- 1 d'anàlisi de la postura dreta i assentada⁷⁰
- 3 d'osteopatia cranial⁴⁷⁻⁴⁹
- 1 de l'articulació temporo-mandibular⁷¹
- 2 de la columna cervical^{72,73}
- 3 de la columna toràcica^{34,74,75}
- 5 de la columna lumbar^{32,33, 76-78}
- 8 de diferents mesures a la pelvis i test de flexió anterior en bipedestació i sedestació^{79,26,80,81,31,61,45,82}
- 1 de mobilitat de l'articulació tibio-peroneo superior (TPS)⁸³
- 1 de dismetria mal·lèols medials²⁷

Avaluació de la qualitat dels articles

No es va identificar cap eina validada i traduïda al català/castellà amb criteris definits per poder avaluar la qualitat dels articles. Si que existeixen articles sobre avaluació qualitativa⁸⁴ o avaluació metodològica⁸⁵, però en cap cas existeixen eines per avaluar la qualitat dels articles.

Existeixen en llengua anglesa dos sistemes de validació que es van definir inicialment al 2003. Un és el nomenat *STARD* (Standards for Reporting of Diagnostic Accuracy) que van definir Bossuyt i altres^{86,87} per articles científics. *STARD* consta de 25 elements per avaluar el rigor científic des del títol, fins la discussió i els resultats, passant per tots els elements que componen un article.

L'altre sistema d'avaluació és el nomenat *QUADAS* (Quality assessment of studies of diagnostic accuracy included in systematic review) que va ser desenvolupat per Whiting⁸⁸ comptant entre els autors amb Bossuyt i que es va centrar més en metodologia de recerca mèdica. Va ser reevaluat al 2006⁸⁹ constatant la seva utilitat. Aquest consta d'un llistat de 14 ítems que es responen amb: si, no, o dubtós (*unclear*).

Vist que a les revisions bibliogràfiques hi ha l'ús dels dos, STARD o QUADAS, i d'algun altre sistema creat pròpiament pels autors, semblava més lògic fer servir el més adient per analitzar les dades tractades i per tant el que més es fa servir en aquestes revisions que és el sistema QUADAS. Ja que aquest sistema *no té traducció validada* en llengua castellana i/o catalana es va decidir per una adaptació per aquest treball. (Taula 1 als annexes). Es tracta en tot cas d'una eina per avaluar la qualitat de precisió d'estudis de diagnòstics.

Extracció de dades

Dels articles seleccionats es van extreure dades dels participants (número, sexe, edat i característiques clíniques), examinadors (número, formació, anys d'experiència), avaluació dels procediments (regió que s'avalua, postura del participant, definició dels tests), resultats estadístics de la fiabilitat inter/intraexaminador (k , CCI, r) i resultat i/o conclusions dels estudis.

Anàlisi estadística

Dels articles seleccionats es van estudiar els resultats estadístics comparant els estudis entre altres de la mateixa regió si era possible. Es van recollir els resultats dels valors kappa (k) comparant-los amb els intervals descrits per Landis & Koch^{90,91} que són els que s'utilitzen com a referència. Hi ha autors que descriuen acceptable la fiabilitat a partir de 0,40⁶⁵ tot i que altres la consideren substancial a partir de 0,60²⁶ i altres defineixen els intervals de manera diferent i descriuen el límit de 0,50 a partir del que podem parlar de fiabilitat moderada¹. Per decidir quina és la fiabilitat "acceptable" es va seguir la descripció original de Landis & Koch de 1977^{90,91} (Taula 2).

Kappa	Nivell d'acord
<0.00	Pobre
0.00-0.20	Lleu
0.21-0.40	Just
0.41-0.60	Moderat
0.61-0.80	Substancial
0.81-1.00	Gairebé perfecte

Taula 2: Mesures d'acord per dades categòriques segons Landis & Koch

Si en algun dels casos no es fes servir el valor de kappa, sinó el valor dels coeficients de correlació intraclasse (CCI) es faria servir la descripció proposta per Hopkins⁹².(Taula 3).

Coefficient de correlació	Descriptor
0.0-0.1	Trivial, molt petita, insubstancial, pràcticament nul
0.1-0.3	Petita, baix, menor
0.3-0.5	Moderada, mitjana
0.5-0.7	Gran, alta, major
0.7-0.9	Molt gran, molt alta, enorme
0.9-1	Prop, pràcticament o gairebé: perfecte, infinita

Taula 3: Descriptors dels coeficients de correlació segons Hopkins

Per a la realització dels gràfics estadístics amb els resultats extrets dels articles es va utilitzar el programa informàtic The Statistical Package for the Social Sciences (SPSS™) versió 20.0 d'IBM.

Planificació de la recerca

Tal com hem indicat anteriorment per realitzar aquest estudi es va diferenciar entre la recerca de textos i articles sobre el tema, i la lectura dels articles més destacats per extraure'n les dades més importants.

Des de la preparació i presentació del protocol d'una recerca finalitzada al desembre del 2010 i fins data de desembre del 2011 es van intentar localitzar el màxim de referències i articles sobre fiabilitat inter/intraexaminador en osteopatia. Es van llegir tan texts provinents de revistes pròpiament d'osteopatia, com d'altres disciplines de teràpia física i/o manual on també es publicuessin textos osteopàtics.

Es van buscar referències sobre paràmetres de fiabilitat i validesa, en textos osteopàtics i d'altres disciplines mèdiques i sanitàries, així com en textos tècnics estadístics i matemàtics.

A partir de novembre de 2011 es va començar a donar forma i definir la inclusió/exclusió d'articles seleccionats i treure conclusions. Es va iniciar el redactat d'aquest text i es van elaborar els resultats i la discussió. (Taula 4)

	Des 2010	Gen 2011	Feb 2011	Mar 2011	Abr 2011	Mai 2011	Jun 2011	Jul 2011	Ago 2011	Set 2011	Oct 2011	Nov 2011	Des 2011	Gen 2012
Preparació versió inicial document	■													
Lectura crítica articles	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Definició "test osteopàtic"		■	■	■										
Definir paràmetres fiabilitat i validesa					■	■	■							
Metodologia Qualitativa							■	■	■					
Anàlisi d'articles: Inclusió/exclusió												■	■	
Anàlisi de Dades: Descripció i relacions												■	■	
Anàlisi Estadística i Conclusions												■	■	
Recerca d'articles	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Entrega Projecte														■

Taula 4: Cronograma de planificació de les tasques del projecte.

Resultats

Dels 183 documents localitzats, es van analitzar exclusivament 25 amb els criteris que consten a la selecció d'articles. Aquests 25 articles van des de l'estudi de la consistència de la postura⁷⁰, fins a l'avaluació dels factors que afecten la fiabilitat de la dismetria dels mal·lèols²⁷.

Tots ells van ser valorats amb els criteris QUADAS com consta la taula 5 als annexes (Quadre de valoració qualitativa segons criteri QUADAS). Els resultats d'aquest sistema de validació van des d'un resultat de cinc ítems positius en un estudi de Cook⁷⁶, i un de Fryer²⁷ fins a un màxim d'onze ítems positius en els casos de Rey Eriz⁷³ i de Snider⁷⁸. Només tres articles tenien menys de 7 ítems positius: els dos de Fryer^{26,27} i el de Cook⁷⁶. La resta tenien 7 o més ítems positius, essent el més comú justament els 7 ítems positius seguit dels 9 ítems positius.

L'ítem que acumulava més negatius és el número 7 que feia referència a la presència o no de prova estàndard de referència, ja que la majoria d'estudis no comparaven els resultats amb radiografies o tomografies computeritzades, o amb cap altra.

Com a resum podem dir que la mitjana d'ítems positius és de $7,96 \pm 1,57$ i una mediana de 8. Podem dir doncs, que el conjunt d'articles escollits per l'estudi tenen una qualitat mínima acceptable.

Per mostrar la informació extreta dels articles inclosos a l'estudi es presenten les dades tabulades i resultats més significatives que es van trobar. (Taula 6 als annexes: Característiques dels estudis inclosos distribuïts per regions i en ordre alfabètic del primer autor).

Del total d'articles seleccionats, 17 dels 25 articles feien servir per presentar els seus resultats el valor k per determinar la fiabilitat. D'aquests 17 tots

presentaven dades de fiabilitat interexaminador, però només 7 d'ells presentaven resultats de fiabilitat intraexaminador.

La resta d'articles presentaven els resultats amb coeficient de correlació intraclasse en el cas de tres d'ells, dos presentaven altres dades com especificitat i sensibilitat, i tres més no presentava cap dada estadística que es podés analitzar.

Agrupant la informació trobada en funció de la regió analitzada es pot concloure que:

L'article sobre la consistència de la postura i els tres articles sobre palpació cranial no presenten cap dada en forma de valor k que permeti comparacions entre ells. Cal dir que l'article de Moran⁴⁷ si presentava dades de fiabilitat inter e intraexaminador però en forma de CCI. Aquest resultats eren d'un resultat de fiabilitat interexaminador de -0,09 a +0,31 que segons els criteris descrits per Hopkins de la Taula 3 són d'insubstancials a petits. En el cas de la fiabilitat intraexaminador per uns dels examinadors era de 0,47 a 0,65 i en l'altre examinador era de 0,52 a 0,73: aquests resultats segons els mateixos criteris és de moderada a alta pel primer, i d'alta a molt alta pel segon.

L'article de validació del test de l'articulació temporo-mandibular (ATM) de García García⁷¹ oferia dades de fiabilitat interexaminador en forma de valor k que eren de 0,534 (0,41 a 0,6) en el test actiu i 0,547 (0,41 a 0,6) en el cas del test passiu essent els dos resultats moderats segons els criteris de Landis & Koch descrits a la Taula 2.

Dels articles de la regió cervical, Rey Eriz⁷³ presentava dades de sensibilitat i valors predictius, però cap de fiabilitat ni inter ni intraexaminador. Si en presentava McPartland⁷² que comparava l'examinació cervical dels punts sensibles de Jones seguint el criteri diagnòstic de disfunció somàtica segons l'acrònim *TART* (tissue texture change – canvis de la textura dels teixits -, abnormal quality of resistance – qualitat anormal de resistència-, restriction of vertebral motion -restricció de la mobilitat vertebral- i tenderness –

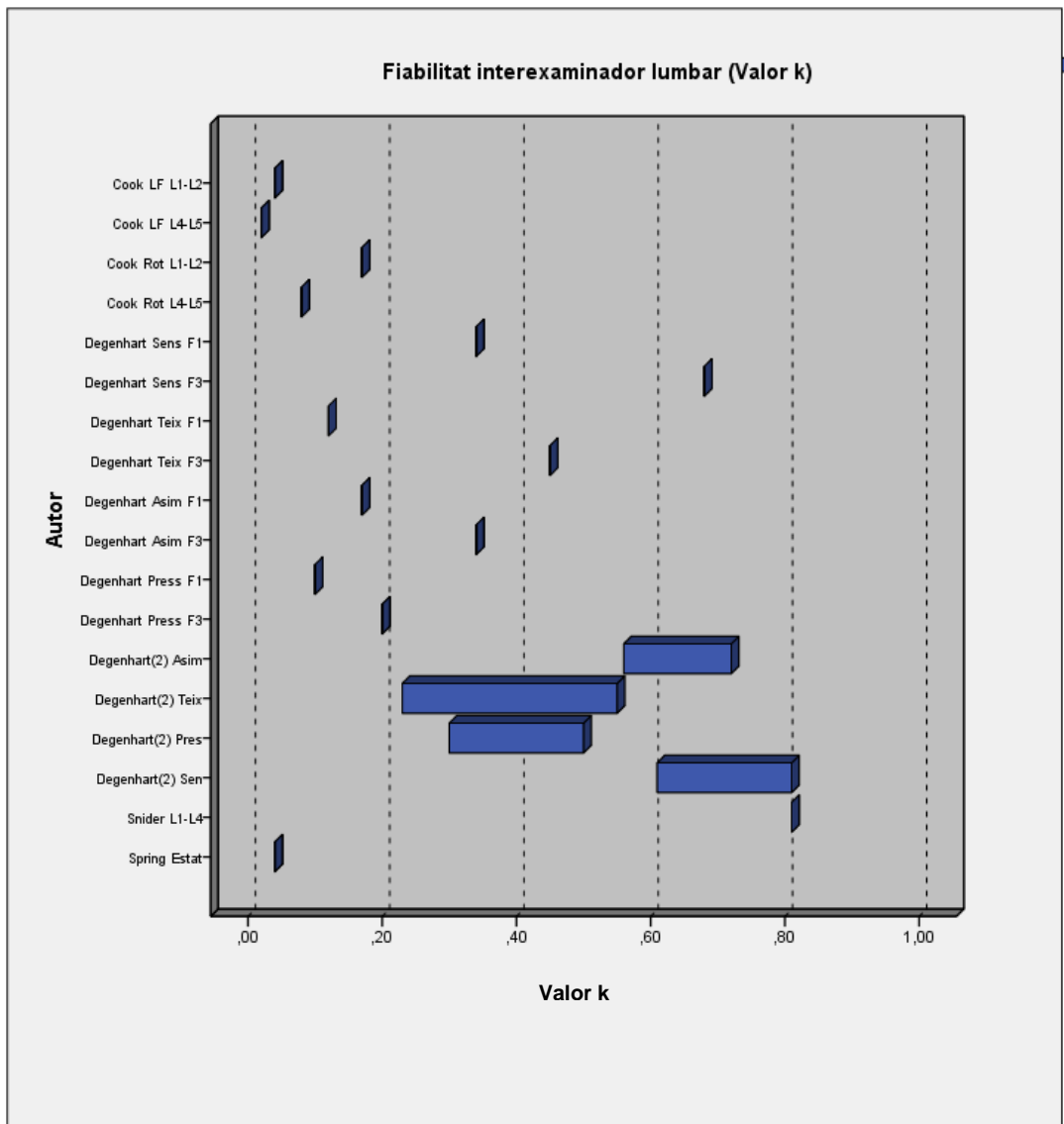
sensibilitat-). Els resultats donaven uns valors de fiabilitat inter-examinador en forma de valor k dels punts sensibles de 0,45 per pacients simptomàtics i 0,19 pels asimptomàtics i pel diagnòstic TART de 0,35 pels simptomàtics i 0,34 pels asimptomàtics. Segons els criteris de Landis & Koch (taula 2) només és moderada la fiabilitat dels punts sensibles en pacients simptomàtics, tot i així es conclou que per pacients asimptomàtics el valor k és més alt en el diagnòstic TART.

Els dos articles de la regió toràcica presentaven resultats de fiabilitat interexaminador. El treball de Ghoukassian³⁴ presentava una fiabilitat en valor k de 0,07, lleu segons Landis & Koch (taula 2) per l'exploració amb el mètode de percussió de Johnson and Friedman.

Paulet⁷⁴ estudiava la fiabilitat de la palpació de anormalitat de la textura dels teixits en la regió toràcica paravertebral amb una fiabilitat de 0,26, justa, i si es calculava només per les cinc primeres palpacions per eliminar la fatiga augmentava però continuava sent justa amb un valor k de 0,32.

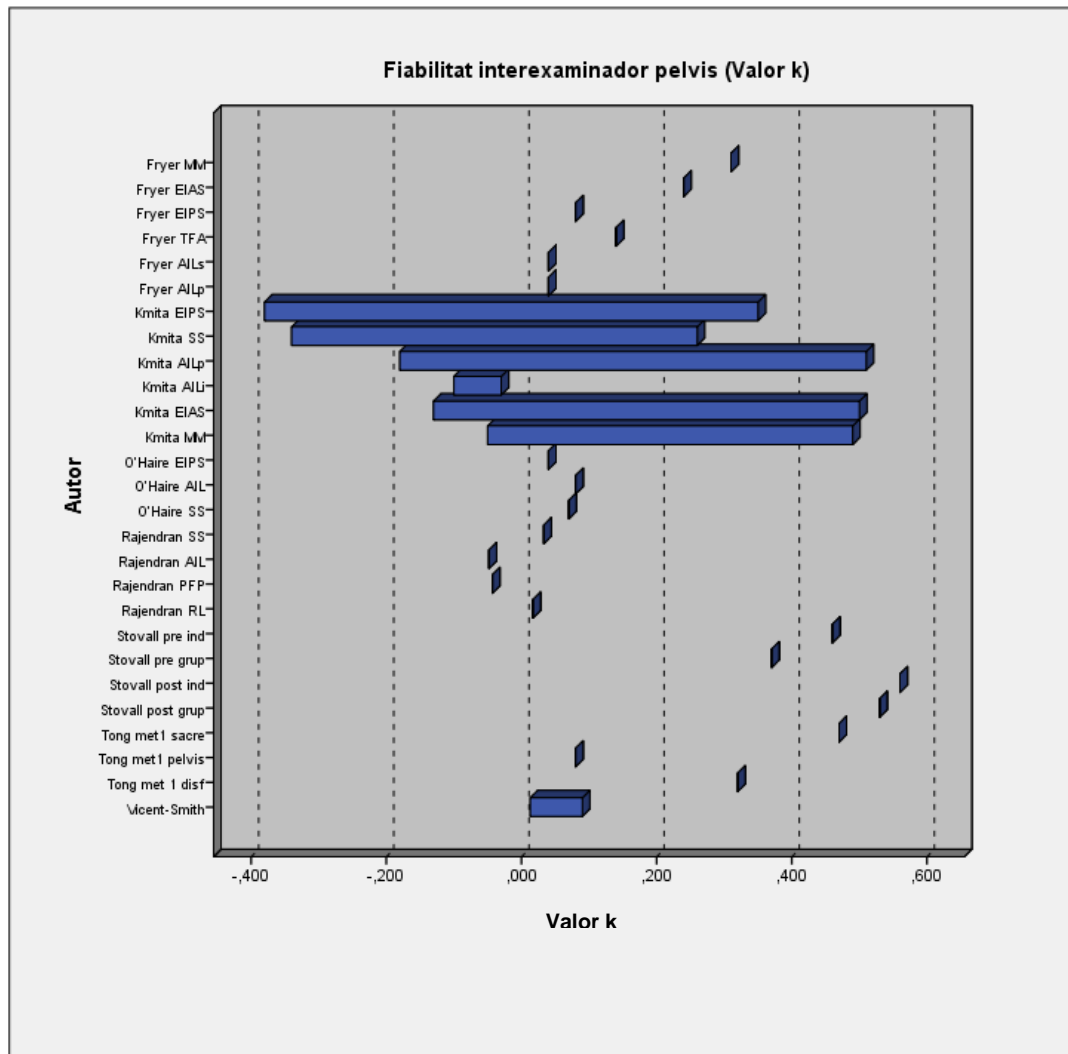
Tots els articles de la regió lumbar presentaven resultats de fiabilitat inter-examinador, però només un d'ells presentava resultats de fiabilitat intraexaminador³³. Els resultats són variables segons l'objectiu de l'estudi i van des de valors k de 0,02 (fiabilitat lleu) en l'estudi de Cook⁷⁶ sobre l'observació de moviments acoblats lumbar a valors k de 0,81 (fiabilitat gairebé perfecte) en el cas de Snider sobre identificació de les espinoses lumbar amb múltiples relleus ossis⁷⁸.

La fiabilitat interexaminador es presenta a mode de gràfic (Gràfic 1). La fiabilitat intraexaminador era només descrita per Spring³³ que estudiava el diagnòstic posicional de la columna lumbar amb un resultat lleu de 0,044 en el valor k.



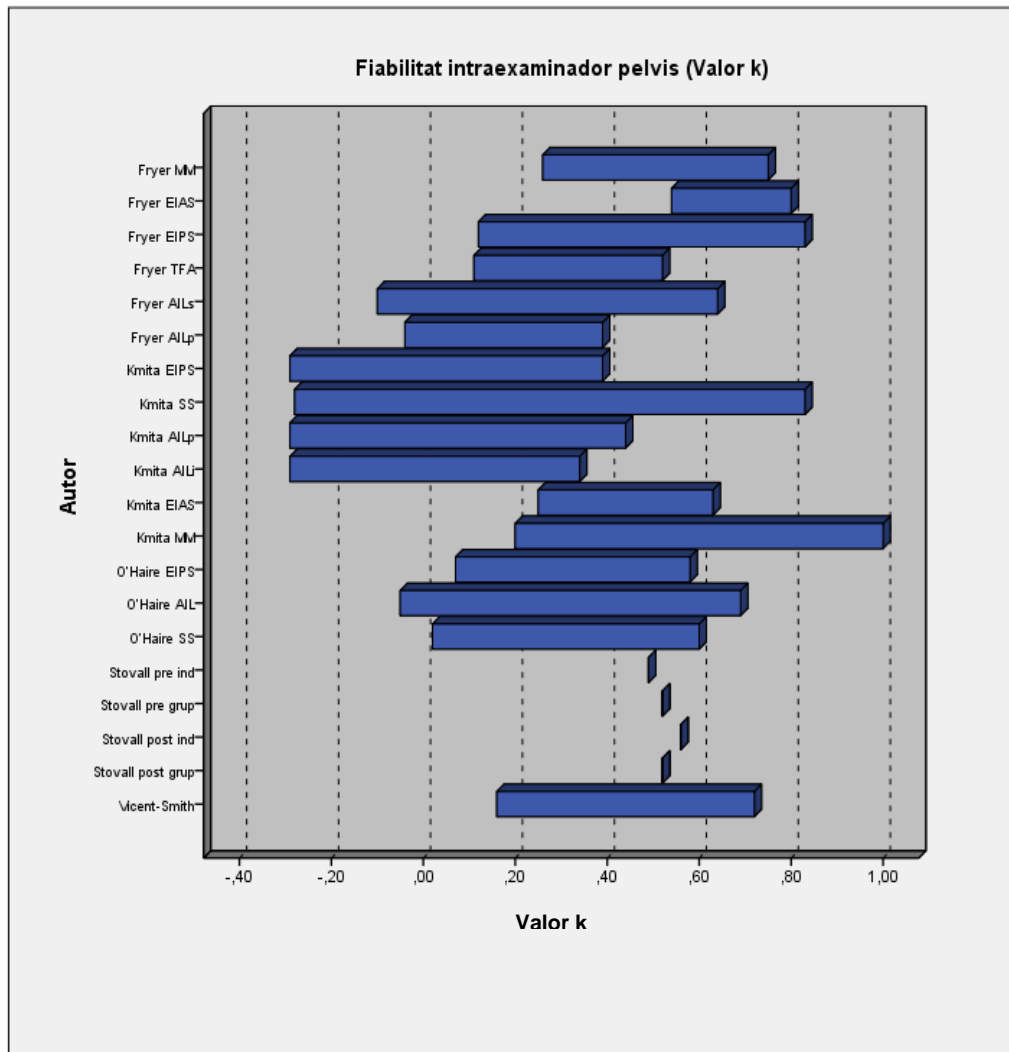
Gràfic 1: Fiabilitat interexaminador lumbar

Dels articles de la regió pèlvica, tots menys el de Bengaard⁷⁹ presentaven resultats en forma de valor k per la fiabilitat interexaminador i d'aquests set, hi havia cinc que presentaven a més dades en forma de valor k per la fiabilitat intraexaminador. En tots els estudis que verifiquen asimetries a nivell del sacre els resultats demostren fiabilitats baixes tant inter com intraexaminador. Es presenten els resultats en forma de gràfic 2: Fiabilitat interexaminador pelvis (Gràfic 2) i gràfic 3: Fiabilitat intraexaminador pelvis (Gràfic 3).



Gràfic 2: Fiabilitat interexaminador pelvis

L'estudi sobre el test de mobilitat de l'articulació tibio-peroneo superior de Bazzolo⁸³ no presentava dades de valor k ni de CCI. Presentava resultats en forma de sensibilitat, especificitat, valor predictiu positiu i valor predictiu negatiu.



Gràfic 3: Fiabilitat intraexaminador pelvis

L'últim estudi era el de Fryer²⁷ que tractava sobre factors que afectaven la fiabilitat intraexaminador i interexaminador de la palpació de la asimetria dels mal·lèols medials en supí. Aquest feia un primer assaig on els resultats eren de 0,22 per la fiabilitat interexaminador i 0,31 (-0,20 a 0,73) per la intraexaminador, justa en els dos casos. Es valoraven els factors que podien afectar aquests resultats i després "d'eliminar-los" la fiabilitat interexaminador passava a ser de 0,94 i la interexaminador també de 0,94 (0,70 a 1), gairebé perfecte en els dos casos segons els criteris de Landis & Koch (taula 2).

Discussió

Tot i la importància que els procediments que es fan servir pel diagnòstic osteopàtic siguin fiables, els resultats que es van obtenir en aquesta revisió bibliogràfica donen com a primer resultat la manca d'estudis realitzats exclusivament des de l'osteopatia. Es van localitzar 183 articles inicials del que només 25 componen l'estudi ja que la resta són d'altres àmbits i camps de la teràpia física i manual com la quiropràxia, la fisioteràpia, etc... L'examen físic i la fiabilitat dels seus tests si aporten gran número de referències, però quan parlem d'osteopatia els articles localitzats és molt baix. Probablement s'ha publicat molt poc perquè molts dels tests que es fan servir ja són validats per altres disciplines.

És interessant destacar que no es van localitzar eines d'avaluació de la qualitat metodològica validades al castellà i/o català, així que seria interessant aquí obrir un nou camí o camp de treball per posteriors revisions bibliogràfiques en aquest idioma.

Tot i això, la qualitat que van demostrar els articles en una escala de 14 ítems era d'una mitjana de 7,96 punts positius, i només tres no arribaven als set ítems. Aquests tres es podrien considerar articles de poca qualitat, però cal destacar que és justament la seva metodologia -que intenta millorar les dificultats dels estudis donant nous paràmetres, refent i revaluant els tests- el que fa que alguns dels ítems qualitius siguin negatius. És el cas de Fryer^{26,27} i els seus dos estudis avaluats.

El primer²⁶ tracta sobre l'efecte que té l'entrenament en la millora de la fiabilitat tant inter com intraexaminador. Un estudi on la mostra va canviant i no tota rep el mateix tipus d'actuació, on no tots els pacients reben el mateix tipus de prova estàndard de referència, i on per intentar aleatoritzar els observadors respecte els subjectes tots aquests són noies, etc, acumularà

molt ítems negatius, tot i que això no sigui equiparable a l' utilitat de les seves conclusions.

L'altre estudi també del mateix autor²⁷ té els mateixos “problemes”. Es tracta d'un estudi per detectar els factors que afecten la fiabilitat també inter e intraexaminador de la dismetria dels mal·lèols. Amb la premissa que dismetries molt petites, de menys de 3 mm, són difícilment valorades com a presents o absents, l'autor després d'un primer intent, cerca un grup exclusivament de noies on la dismetria sigui segons ell probable, amb l'objectiu que els examinadors tornin a valorar en aquest grup l'existència o no de la dismetria. Com en l'altre cas, ni tota la mostra rep la mateixa intervenció, ni l'espectre de pacients era representatiu, ni la prova era independent de l'estàndard de referència. L'objectiu no és tant validar el test de la palpació de la dismetria, sinó concloure quins factors la poden afectar i potser per això dins uns barems tancats el resultat qualitatiu és baix, però les interpretacions i conclusions de l'estudi són molt valuoses.

Cal destacar també que pel tipus d'estudis escollits, en pocs es comparava amb un estàndard de referència els test estudiats, ja que l'objectiu de la majoria d'estudis no es validar el propi test. Per contra si ho fan amb èxit García García⁹³ i Rey Eriz⁷³, que són capaços de validar els seus tests amb nivells de 97% i 72,7% de sensibilitat i especificitat pel primer i sensibilitat d'un 80 a un 100% al segon, sinó estudiar la fiabilitat dels tests estudiats intra, però sobretot interexaminador.

Respecte a la metodologia dels estudis, n'hi havia estudis on la mostra constava d'un nou “model pèlvic” dissenyat per entrenar la precisió de la fiabilitat de la palpació com és l'estudi de Stovall⁹⁴, fins a estudis on era molt escassa, de 3 pacients totals en el cas de Cook⁷⁶ fins a mostres tant àmplies com 734 subjectes que a la vegada eren els 734 observadors en el cas de l'estudi de Sergueer⁴⁹. La majoria d'estudis, deu del total, tenien entre 9 i 11 subjectes de mostra, i només 6 estudis superaven els 40 pacients de mostra. Tot i que existeixen càlculs per determinar el número mínim segons el nombre d'observadors, sembla que la majoria de les mostres eren massa

petites per què les conclusions i anàlisis siguin estadísticament significatives. A més hi ha mostres on tots el subjectes són d'un mateix gènere, d'altres on tots els subjectes són estudiants universitaris, i d'altres on tots són voluntaris "sans". Que siguin o no simptomàtics, no exclou que hi hagi subjectes que tot i no presentar clínica tinguin disfuncions somàtiques, però si que caldria que les mostres fossin més heterogènies i per tant més semblants a la realitat de la pràctica diària.

Pel que fa al número d'observadors era també variable en mida i en experiència. Hi ha 13 estudis on intervenien estudiants d'osteopatia, altres on eren estudiants junt amb osteòpates amb experiència i altres només d'osteòpates amb experiència.

Tampoc era una dada constant el fet que hi hagi o no entrenament previ, sent 11 dels estudis els que especificaven que havia existit entrenament previ, que normalment era d'una hora, que es feia el dia de l'assaig clínic. Podrien semblar dades irrellevants, però importants a l'hora de comparar els resultats dels estudis, sobretot si hi ha moltes diferències entre ells. Potser caldria consensuar i estandarditzar aquest tipus de dades amb l'objectiu de fer comparables els estudis i les dades contingudes en ells.

Pel que fa als objectius de la revisió es van poder concloure diferents resultats: podem dir que els tests osteopàtics són aquells que intenten determinar la presència d'una disfunció somàtica en els individus valorats. La disfunció somàtica es descriu pels criteris diagnòstics de l'acrònim STAR (o TART) que correspon a:

- Sensibilitat alterada de la zona a la palpació o pressió. No és sinònim de dolor. (S)
- Anormalitat en la textura dels teixits que es poden palpar. (T)
- Asimetria de la forma, la posició, del moviment,...(A)
- Restricció del moviment dins de l'arc fisiològic on apareix una barrera restrictiva. (R)

Tots els articles estudiats expliquen que, a partir dels criteris diagnòstics de la descripció de disfunció somàtica, valoren un d'ells per portar a terme el seu estudi, com en el cas de la pelvis les asimetries del sacre o els relleus ossis, o bé estudien algun d'aquests criteris com objectiu del propi estudi a partir de la descripció.

En el cas de la postura⁷⁰ estudien si els canvis dels angles i mesures preses es modifica o no, per tant de la simetria/asimetria.

En el cas dels estudis en el camp cranial⁴⁷⁻⁴⁹ s'estudia la presència del moviment respiratori primari (MRP) i la quantitat i qualitat d'aquest, per tant la presència del moviment.

En el cas de l'ATM⁷¹ la mateixa autora ja defineix que alteracions en la simetria i facilitats o dificultats de moviment són segons el seu criteri aspectes per determinar la presència de disfunció en aquella articulació.

En els casos de la columna cervical el primer dels estudis⁷² ja diferenciava entre diagnòstic per punts sensibles de Jones o de criteris de disfunció somàtica segons STAR. El segon⁷³ dels estudis cervicals cercava zones hipomòbils cervicals, i per tant restriccions del moviment.

En els casos de la columna toràctica, el primer³⁴ valorava la columna toràctica mitjançant una percussió i la resposta com a alteració dels teixits. I el segon⁷⁴ cercava directament zones on hi ha anormalitat dels teixits paravertebrals.

L'article sobre identificació de zones en disfunció de la columna lumbar i toràctica⁷⁵ fa servir tests actius i passius per determinar zones d'hipomobilitat, per tant de restricció de la mobilitat.

En els estudis de la regió lumbar, excepte en el de Snider⁷⁸ que té per objectiu localitzar l'apòfisi espinosa, a tota la resta o l'avaluació és amb tot el criteri STAR i així ho especifica, per exemple Degenhart³²; o bé són criteris de restricció de mobilitat i/o anormalitats dels teixits.

A la pelvis, la majoria dels tests (set dels vuit) tracten de fer un diagnòstic posicional de la pelvis basats en el model teòric de Fred L. Mitchell. Es busquen asimetries de les espines postero superiors de la pelvis, com del solc sacre o de l'angle infero lateral d'aquest. Només Vicent-Smith⁸² realitza una avaluació de la fiabilitat inter/intraexaminador del test de flexió anterior, que denota un costat amb restricció de moviment.

Els dos darrers articles de la regió dels turmells segueixen aquest patró de valorar restricció de moviment⁸³ o dismetria, i per tant una asimetria²⁷.

Per tant els tests osteopàtics anirien a la recerca d'aspectes més qualitius que no pas quantitius, aspectes que no es busquen en altre tipus de tests com els ortopèdics⁶⁷, o com en els tests de provocació de dolor dels que hi ha tants en la literatura científica^{95,96}. La diferencia és intentar definir àrees que pels seus canvis estiguin en disfunció, encara que no presentin dolor o simptomatologia. Una àrea en disfunció donarà lloc a una funció alterada del sistema musculoesquelètic i per tant com diu Fryer²⁶ disfuncions anatòmiques, i per tant viscerals.

El següent objectiu i potser més important alhora de realitzar aquest treball era el referent a la fiabilitat dels tests. Com ja s'ha comentat la fiabilitat fa referència a la reproductibilitat de les troballes, no a la precisió dels tests. Els tests han de ser fiables i han de poder ser reproduïbles tant per un mateix com per altres terapeutes.

La fiabilitat intraexaminador mesura l'habilitat del terapeuta d'obtenir el mateix resultat quan s'avalua de forma repetida al mateix pacient. Aquest tipus de fiabilitat és la més discutida ja que hi ha autors com Degenhart³² o Haas⁹⁷ que parlen de la dificultat de cegar l'examinador entre observacions. Fins i tot Haas⁹⁷ diu que ha de tenir més pes la fiabilitat interexaminador que la intra quan es volen elaborar conclusions sobre fiabilitat. Probablement per aquesta raó tan pocs estudis tenen com a part dels seus objectius la fiabilitat intraexaminador.

De totes maneres també hi ha discrepàncies sobre els resultats dels estudis on es calcula aquest valor de fiabilitat. Kmita⁸⁰ diu en el seu estudi que la fiabilitat interexaminador és generalment més baixa que la intraexaminador. En canvi Spring³³ diu que la fiabilitat intraexaminador no és més gran que la interexaminador al seu estudi, al contrari que molts tests de palpació previs. El cas és que l'estudi de Spring³³ és l'únic localitzat realitzat per osteòpates que es calcula la fiabilitat intraexaminador a la regió lumbar, ja que la resta d'estudis amb dades de fiabilitat intraexaminador són tots de la regió de la pelvis. I així sí la fiabilitat interexaminador de la pelvis dona resultats més aviat pobres, els de la zona lumbar són més moderats. Potser el fet que la fiabilitat lumbar sigui millor que la pèlvica pot modificar d'alguna manera la fiabilitat intraexaminador, però no es disposen de suficients dades com per arribar a aquesta conclusió.

De totes formes dels resultats de la revisió podem deduir que la fiabilitat intraexaminador de la pelvis en conjunt és justa i globalment no arriba al 0,4. Si eliminéssim dels resultats tots aquells en què ja hi ha valors mínims de fiabilitat intraexaminador negatius i que corresponen a les referències del sacre, tant del solc com de l'angle infero-lateral, la fiabilitat milloraria fins a límits de moderada. La fiabilitat del test de flexió anterior en bipedestació només s'avalua en l'estudi de Vicent-Smith⁸² i amb uns resultats de 0,46, una fiabilitat moderada que per alguns autors ja es clínicament vàlida. El test de flexió anterior en sedestació s'estudia, junt amb altres punts anatòmics, en un dels estudis de Fryer²⁶ amb un resultat de 0,41, també fiabilitat moderada.

La fiabilitat intraexaminador de la regió de la pelvis més elevada de manera aïllada és la dels mal·lèols que té un valor mig de 0,64 -una fiabilitat substancial- seguida per la fiabilitat de les espines ilíaques antero-superior (EIAS) que es de 0,54 -moderada- i les espines ilíaques postero-superiors (EIPS) amb un valor de 0,30 -fiabilitat justa -.

Amb els resultats obtinguts no podem dir que en el seu conjunt la fiabilitat intraexaminador dels tests avaluats sigui acceptable.

Pel que fa als resultats de la fiabilitat interexaminador, primer de tot podem dir que és un dels objectius principals dels tests en els articles que s'han localitzat en aquesta recerca. Tots els estudis de la columna lumbar calculaven la fiabilitat interexaminador, i tots menys un, que si ho indica al text⁷⁹, de la pelvis ho fan també.

Respecte a la columna lumbar hi ha autors com Seffinger⁹⁸ o Stochkendahl⁵¹ que en diferents revisions arribaven a la conclusió que els tests de provocació de dolor eren fiables tan intra com interexaminador; els tests de mobilitat i de localització de relleus ossis eren més fiables intraexaminador que interexaminador; i que altres tests diferents de provocació de dolor tenien poca fiabilitat interexaminador. Pocs estudis, com diu Spring³³, tenen com objectiu investigar la palpació estàtica com una eina de diagnòstic, i els resultats dels estudis existents són molt variables.

A més, molts dels autors com Degenhardt^{32,77} o d'altres argumenten que la fiabilitat interexaminador té algunes limitacions importants com és que el cos humà no és estàtic, sinó dinàmic i el sistema neuromusculoesquelètic canvia de nivell cada segon depenent dels impulsos o de les experiències individuals, per tant és impossible palpar dos vegades el mateix. Si afegim una de les altres limitacions que s'esmenten, com és que al realitzar els tests diferents vegades sobre la mateixa persona ja s'està modificant el resultat, ja que es modifiquen les condicions dels subjectes.

En la revisió que es va efectuar els resultats demostraven que en el seu conjunt els tests osteopàtics a la pelvis tenien un valor de 0,31, sent una fiabilitat justa. Cal destacar aquí els estudis de Degenhardt³², on al primer es realitza una investigació en tres fases. A la primera fase es testen tota una sèrie de test palpatoris a doble cec, i aquells que donen un resultat k menor de 0,40 passen a una segona fase on tots els observadors fan sessions d'entrenament fins que consensuen entre tots l'acord de què i com s'ha de realitzar el test. Quan arriben a un alt nivell d'acord, passen a una tercera fase on es torna fer un estudi a doble cec. En aquest estudi s'arriba a la conclusió que l'entrenament de consens pot millorar la fiabilitat

interobservador de test palpatoris osteopàtics de la columna lumbar. A més dos d'aquests milloren fins a nivells acceptables de valor k com és el cas de la palpació dels canvis als teixits, i de la sensibilitat.

En el segon estudi de Degenhardt⁷⁷ es valora el canvi de fiabilitat interexaminador de tests palpatoris durant un període de quatre mesos, on s'arribava a la conclusió que els examinadors eren capaços de mantenir, i fins i tot millorar la fiabilitat interexaminador de quatre tests de diagnòstic palpatori lumbar.

En l'estudi de Snider⁷⁸ sobre la precisió de la palpació de les espinoses de la primera a la quarta vertebra lumbar es va arribar també a la conclusió de que a part d'un alt nivell d'identificació de les espinoses, la fiabilitat depenia de molts aspectes, entre ells el nivell d'experiència dels examinadors. Hi ha altres estudis⁹⁹ des d'altres disciplines de la teràpia manual on l'objectiu era verificar si existia diferència entre les habilitats palpatories entre terapeutes amb i sense experiència, on s'arribava a la conclusió que l'experiència no millorava en si mateixa les capacitats per aprendre.

Amb tot això, tot i que l'experiència, l'entrenament i el consens siguin eines útils en un futur per millorar la fiabilitat dels tests osteopàtic lumbar continuen tenint una fiabilitat justa pel seu ús clínic.

Per últim, la fiabilitat interexaminador dels tests de la pelvis és la que més estudis acumula. I és la que ofereix un valor de fiabilitat en el seu conjunt més baix.

D'estudis sobre fiabilitat interexaminador a la regió pèlvica hi ha molts des de totes les disciplines de la teràpia manual i molt tipus de test. Hi ha autors com Freuberg¹⁰⁰ que els separen en tres grans grups: localització de relleus ossis i asimetries, mobilitat de determinats punts anatòmics activa o passivament, i per últim, test de reproducció del dolor.

En diferents revisions bibliogràfiques com la de Haneline⁵⁵ s'arribava a la conclusió que els índexs d'acord dels tests era generalment baixos i que

molts dels tests de provocació de dolor tenen nivells acceptables de valor k. També arriben a aquesta conclusió Riddle¹⁰¹ i Robinson⁶³ entre d'altres. També com diu Vicent-Smith⁸² és millor la combinació de test que no un de sol o aïllat. De totes formes cal esmentar diversos problemes que es van repetint en moltes de les investigacions. Un d'ells és com diu O'Haire⁸¹ l'error que hi pot haver en la localització dels relleus ossis.

El resultat obtingut d'aquesta revisió, és que la fiabilitat interexaminador dels tests osteopàtics a la regió pèlvica és justa amb un valor mig de $k=0,21$. Només l'estudi de Stovall⁹⁴ en que l'objectiu era construir un model pèlvic artificial per millorar la fiabilitat i precisió de la palpació, arribava a nivells de valor k de 0,40 i per tant moderats. La resta d'estudis no arribava a aquests nivells i si eliminem els resultats de Stovall⁹⁴ la fiabilitat passa a ser lleu. Això concorda amb els resultats obtinguts en estudis previs.

Així podem dir que la fiabilitat interexaminador dels tests osteopàtics per la pelvis té uns valors justos que no els fan acceptables pel seu ús clínic.

En conjunt es pot concloure que cap de les fiabilitats estudiades té nivells de fiabilitat que aconsellin el seu ús. Als estudis de fiabilitat sobre els test diagnòstics són els investigadors, els que realitzen les proves, qui són l'objecte d'investigació pròpiament, no el subjecte passiu, i per tant són un dels ítems que s'està avaluant. Caldria doncs intentar arribar a un consens entre la validesa i la fiabilitat dels tests per tal de saber treure el màxim d'informació sobre ells. I caldria determinar amb molta claredat durant l'aprenentatge dels tests que es busca, perquè es busca i com es busca. És necessari un acord per què tothom cerqui el mateix i de la mateixa manera.

A més, si com molts autors diuen que els test realitzats sobre un cos dinàmic i en constant canvi no poden ser idèntics en dos avaluacions diferents, afegim que la disfunció somàtica és inestable, inconsistent i modificable, com es pot trobar, com es pot fixar i com es pot determinar la seva correcció?

Ja hi ha autors¹⁰² que proposen que abandonar els mètodes de palpació perquè són poc fiables per altres més sofisticats seria un memorable error, i

recomanen combinar diferents tècniques. Es qüestionen si la palpació és una qüestió de ciència o un art. Potser com diuen el problema de la fiabilitat dels tests no està en els tests, sinó en com es desenvolupa la fiabilitat, la sensibilitat i els tests vàlids.

Potser cal tenir en compte que actualment es considera el model de la medicina basada en l'evidència com l'únic vàlid, tot i que existeixen autors¹⁰³ que proposen altres dissenys més qualitius i menys pragmàtics. Sembla que davant les actuals circumstàncies socials i científiques, els professionals sanitaris s'hagin basat en l'evidència científica fent ús de la metodologia quantitativa com a principal procediment per resoldre problemes clínics. Aquesta corrent està basada en models probabilístics que fa servir de manera conscient i explícita la millor evidència científica disponible per prendre decisions sobre l'atenció als pacients.

Si es considera la medicina com una ciència social que es nodreix també de la psicologia, la pedagogia i la sociologia, sembla que el punt de vista quantitatiu necessitaria complementar-se d'un més qualitatiu. Com diuen els autors, la percepció de salut o malaltia d'un pacient no depèn únicament de la capacitació tècnica del facultatiu, sinó dels vincles i xarxes, la capacitat d'escolta, la sensibilitat, la paciència i la receptivitat d'aquest. A més recorden que s'ha incidit tant en el diagnòstic i el tractament que s'ha oblidat el tenir cura de la persona. Seguint aquest exemple, potser s'estan fent esforços per validar l'osteopatia amb uns procediments en que s'obliden aspectes qualitius, i volem quantificar coses difícils de quantificar científicament.

Dins d'aquesta corrent podríem incloure Irvin Korr¹⁰⁴ que ja al 1991, reflexionava sobre la necessitat de recerca i el canvi necessari de paradigma. Proposava projectes rellevants per la teoria i pràctica osteopàtica i per desenvolupar protocols vàlids, i dissenys alternatius en consonància amb la filosofia osteopàtica.

Per últim afegir com a limitacions d'aquest estudi la falta d'articles de tests osteopàtics analitzats i la variabilitat d'aquests. Existeixen molts estudis des de diverses disciplines que ens són útils, sobretot en l'ús de test diagnòstics de la regió pèlvica, però pocs han estat realitzats per osteòpates. Menys encara realitzats en castellà i/o català, tot i destacar l'esforç d'algunes publicacions com "Osteopatía Científica" i entintats¹⁰⁵ que promouen la recerca en osteopatía. No podem deixar d'aconsellar que es continuï investigant tant en el referent als tests, com als propis procediments que s'utilitzen. Només així s'arribaran a tenir suficients resultats com per poder tenir criteris per avaluar i corroborar l'osteopatía com a ciència.

Conclusió

Les conclusions d'aquesta revisió literària són que no hi ha cap test osteopàtic dels estudiats que tingui uns valors acceptables de fiabilitat. La fiabilitat intraexaminador és més alta que la interexaminador però totes amb nivells de valor k baixos. La fiabilitat intraexaminador a les lumbar és d'un nivell moderat i a la pelvis justa.

Pel que fa a la fiabilitat interexaminador, ni a la pelvis ni a la regió lumbar superen el nivell de justa, essent més alta entre les dues, la fiabilitat interexaminador lumbar.

Amb això no podem dir que l'ús clínic dels test osteopàtics estudiats sigui fiable.

Remarcar que no existeixen eines d'avaluació qualitativa validades al català/castellà.

Bibliografia

1. Cleland J, Cleland J. Netter. Exploración clínica en ortopedia: un enfoque para fisioterapeutas basado en la evidencia. 13th ed. Elsevier , editor. Barcelona: Masson; 2006.
2. Bruton A, Conway JH, Holgate ST. Reliability: What is it, and how is it measured? *Physiotherapy*. 2000 Febrer; 86(2): 94-99.
3. Lucas N, Bogduk N. Diagnostic reliability in osteopathic medicine. *Int J Osteopath Med*. 2011 Juny; 14(2): 43-47.
4. Batterham AM, George KP. Reliability in evidence-based clinical practice: a primer for allied health professional. *Physical Therapy in Sport*. 2003 Agost; 4(3): 122-128.
5. Weir JP. Quantifying Test-Retest Reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res*. 2005 Febrer; 19(1): 231-240.
6. Abbott JH, Mercer SR. Lumbar segmental hipomobility: criterion-related validity of clinical examination items (a pilot study). *NZ Journal of phisiotherapy*. 2003 Març; 31(1): 3-9.
7. Cortés Vega MD, Rodríguez Blanco C, González García P, Ribelles Mayo J, Oliva Pascual-Vaca Á. Análisis de la fiabilidad palpatoria del pulso en la arteria radial mediante sonografía Doppler. *Osteopatía científica*. 2009; 4(3): 93-103.
8. Cuthbert SC, Goodheart Jr GJ. On the reliability and validity of manual muscle testing: a literature review. *Chiropr Osteopat*. 2007 Març; 15(4).
9. Gnat R, Kuszewski M, Koczar R, Dziewonska A. Reliability of the passive knee flexion and extension tests in healthy subjects. *J Manipulative Physiol Ther*. 2010 Nov-Des; 33(9): 659-665.
10. Divya Bharatkumar A, Bussey MD, Mani R, Jayakaran P, Aldabe D, Milosavljevic S. Inter-tester reliability of non-invasive technique for measurement of innominate. *Man Ther*. 2011 En premsa.
11. Johansson F. Interexaminer reliability of lumbar segmental mobility tests. *Man Ther*. 2006; 11: 331-336.
12. Hestbæk L, Leboeuf-Yde C. Are Chiropractic Tests for the Lumbo-Pelvic Spine Reliable and Valid? A Systematic Critical Literature Review. *J Manipulative Physiol Ther*. 2000 Maig; 23(4): 258-275.
13. Potter NA, Rothstein JM. Intertester Reliability for Selected Clinical Tests of the

Sacroiliac Joint. *Phys Ther.* 1985 Novembre; 65(11): 1671-1675.

14. Pool JJ, Hoving JL, de Vet HC, van Mameren H, Bouter LM. The interexaminer reproducibility of physical examination of the cervical spine. *J Manipulative Physiol Ther.* 2004 Febrer; 27(2): 84-90.
15. Otoni do Carmo Carvalhais V, Lara de Araújo V, Rezende Souza T, Gomes Pavan Gonçalves G, de Melo Ocarino J, Teixeira Fonseca S. Validity and reliability of clinical tests for assessing hip passive stiffness. *Man Ther.* 2011 Juny; 16(3): 240-245.
16. Gimpel JR, Boulet JR, Errichetti AM. Evaluating the Clinical Skills of Osteopathic Medical Students. *J Am Osteopath Assoc.* 2003 Juny; 103(6): 267-279.
17. Evans DW, Foster NE, Underwood M, Vogel S, Breen AC, Pincus T. Testing the effectiveness of an innovative information package on practitioner reported behaviour and beliefs: The UK Chiropractors, Osteopaths and Musculoskeletal Physiotherapists Low back pain Management (COMPLeMENT) trial. 2005 Juliol. Disponible a <http://www.biomedcentral.com/1471-2474/6/41>.
18. Fryer G, Johnson JC, Fossum C. The use of espinal and sacroiliac joint procedures within the British osteopathic profession. Part 1: Assessment. *Int J Osteopath Med.* 2010 Desembre; 13(4): 143-151.
19. French SD, Green S, Forbes A. Reliability of Chiropractic methods commonly used to detect manipulable lesions in patients with chronic low-back pain. *J Manipulative Physiol Ther.* 2000 Maig; 23(4): 231-238.
20. London S. The assessment of clinical practice in osteopathic education: Is there a need to define a gold standard? *Int J Osteopath Med.* 2008 Desembre; 11(4): 132-136.
21. Lewis LK, Williams MT, Olds TS. Development and psychometric testing of an instrument to evaluate cognitive skills of evidence based practice in student health professionals. *BMC Med Educ* (publicació online). 2011 Octubre; 11(77): <http://www.biomedcentral.com/1472-6920/11/77>.
22. van der Wurff P, Buijs EJ, Groen GJ. A multitest regimen of pain provocation tests as an aid to reduce unnecessary minimally invasive sacroiliac joint procedures. *Arch Phys Med Rehabil.* 2006 Gener; 87: 10-14.
23. van der Wurff P, Meyne W, Hagmeijer RHH. Clinical tests of the sacroiliac joint. A systematic methodological review. Part 2: Validity. *Man Ther.* 2000 Maig; 5(2): 89-96.

24. van der Wurff P, Hagmeijer RHM, Meyne W. Clinical tests of the sacroiliac joint. A systematic methodological review. Part 1: Reliability. *Mant Ther.* 2000 Febrer; 5(1): 30-36.
25. O'Keefe PM, Fryer G. Inter-examiner and intra-examiner reliability of the seated flexion test. School of Health Sciences, Victoria University..
26. Fryer G, McPherson HC, O'Keefe P. The effect of training on the inter-examiner and intra-examiner reliability of the seated flexion test and assessment of pelvic anatomical landmarks with palpation. *Int J Osteopath Med.* 2005 Agost; 8: 131-138.
27. Fryer G. Factors affecting the intra-examiner and inter-examiner reliability of palpation for supine medial malleoli asymmetry. *Int J Osteopath Med.* 2006; 9: 58-65.
28. Fryer G, Morse CM, Johnson JC. Spinal and sacroiliac assessment and treatment techniques used by osteopathic physicians in the United States. *Osteopath Med Prim Care.* 2009 Abril; 3(4).
29. Laslett M, Aprill CN, McDonald B, Young SB. Diagnosis of Sacroiliac Joint Pain: Validity of individual provocation. *Man Ther.* 2005 Agost; 10(3): 207-218.
30. Laslett M. Evidence-Based Diagnosis and Treatment of the Painful Sacroiliac Joint. *J Man Manip Ther.* 2008; 16(3): 142-152.
31. Rajendran D, Gallagher D. The assessment of pelvic landmarks using palpation: A reliability study of undergraduate students. *Int J Osteopath Med.* 2011 Juny; 14(2): 57-60.
32. Degenhardt BF, Snider KT, Snider EJ, Johnson JC. Interobserver Reliability of Osteopathic Palpatory Diagnostic Tests of the Lumbar Spine: Improvements From Consensus Training. *J Am Osteopath Assoc.* 2005 Octubre; 105(10): 465-473.
33. Spring F, Gibbons P, Tehan P. Intra-examiner and inter-examiner reliability of a positional diagnostic screen for the lumbar spine. *Journal of Osteopathic Medicine.* 2001; 4(2): 47-55.
34. Ghoukassian M, Nicholls B, McLaughlin P. Inter-examiner reliability of the Johnson and Friedman percussion scan of the thoracic spine. *J Osteopath Med.* 2001 Abril; 4(1): 15-20.
35. Callaghan JP, Dunk N. The reliability of quantifying upright standing postures as a baseline diagnostic clinical tool. *J manipulac Physiol Ther.* 2004 Octubre; 27(8): 534-535.

36. Dunk NM, Chung YY, Compton DS, Callaghan JP. The reliability of quantifying upright standing postures as a baseline diagnostic clinical tool. *J Manipulative Physiol Ther.* 2004 Febrer; 27(2): 91-96.
37. Dunk NM, Lalonde J, Callaghan JP. Implications for the use of postural analysis as a clinical diagnostic tool: reliability of quantifying upright standing spinal postures from photographic images. *J Manipulative Physiol Ther.* 2005 Juliol-Agost; 28(6): 386-392.
38. Moran R, Ljubotenski E. Reliability of visual assessment for lumbar curves in standing: Does clinical experience matter? *Int J Osteopath Med.* 2006 Març; 9(1): 42.
39. Cadogan A, Laslett M, Hing W, McNair P, Williams M. Interexaminer reliability of orthopaedic special tests used in the assessment of shoulder pain. *Man Ther.* 2011 Abril; 16(2): 131-135.
40. Esparza Ferrera C, Ricard F. Estudio comparativo de pruebas ortopédicas para la valoración y confirmación diagnóstica de tendinopatía del músculo supraespinoso. *Osteopatía científica.* 2009; 4(3): 86-92.
41. Hollerwöger D. Methodological quality and outcomes of studies addressing. *Man Ther.* 2006; 11: 93-98.
42. Lucas NE. Motion Palpation: It's time to accept the evidence. *J Manipulative Physiol Ther.* 2000 Gener; 23(1): 60.
43. McGrath MC. Palpation of the sacroiliac joint: An anatomical and sensory challenge. *Int J Osteopath Med.* 2006 Setembre; 9(3): 103-107.
44. Meoli FG, Wallace WS, Kaiser-Smith J, Shen L. Relationship of osteopathic medical licensure examinations with undergraduate admission measures and predictive value of identifying future performance in osteopathic principles and practice/osteopathic manipulative medicine courses and rotations. *JAOA.* 2002 Novembre; 102(11): 615-620.
45. Tong HC, Heyman OG, A. LD, Isser MM. Interexaminer Reliability of Three Methods of Combining Test Results to Determine Side of Sacral Restriction, Sacral Base Position, and Innominate Bone Position. *J Am Osteopath Assoc.* 2006 Agost; 106(8): 464-468.
46. Van Genderen FR, De Bie RA, Helders PJM, Van Meeteren NLU. Reliability Research: Towards a More Clinically Relevant Approach. *Phys Ther Rev.* 2003; 8(4): 169-176.
47. Moran RW, Gibbons P. Intraexaminer and Interexaminer Reliability for Palpation of the Cranial Rhythmic Impulse at the Head and Sacrum. *J*

Manipulative Physiol Ther. 2001 Març/Abril; 24(3): 183-190.

48. Sommerfeld P, Kaider A, Klein P. Inter- and intraexaminer reliability in palpation of the “primary respiratory mechanism” within the “cranial concept”. *Man Ther.* 2004; 9: 22-29.
49. Sergueef N, Greer MA, Nelson KE, Glonek T. The palpated cranial rhythmic impulse (CRI): Its normative rate and examiner experience. *Int J Osteopath Med.* 2011 Març; 14(1): 10-16.
50. Meijne W, van Neerbos K, Aufdemkampe G, van der Wurff P. Intraexaminer and interexaminer Reliability of the Gillet Test. *J Manipulative Physiol Ther.* 1999 Gener; 22(1): 4-9.
51. Stochkendahl MJ, Christensen HW, Hartvigsen J, Vach W, Haas M, Hestbaek L, et al. Manual Examination of the Spine: A Systematic Critical Literature Review of Reproducibility. *J Manipulative Physiol Ther.* 2006 Juliol/Agost; 29(6): 475-495.
52. Szadek KM, van der Wurff P, van Tulder MW, Zuurmond WW, Perez RSGM. Diagnostic validity of criteria for sacroiliac joint pain: A systematic review. *The journal of pain.* 2009 Abril; 10(4): 354-368.
53. May S, Littlewood C, Bishop A. Reliability of procedures used in the physical examination of non-specific low back pain: A systematic review. *Aust J Physiother.* 2006; 52: 91-102.
54. Vaughan B. Inter-examiner reliability in detecting cervical spine dysfunction: A short review. *Journal of Osteopathic Medicine.* 2002; 5(1): 24-27.
55. Haneline MT, Young M. A Review of Intraexaminer and Interexaminer Reliability of Static Spinal Palpation: A Literature Synthesis. *J Manipulative Physiol Ther.* 2009 Juny; 32(5): 379-386.
56. van Trijffell E, Anderegg Q, Bossuyt PMM, Lucas C. Inter-examiner reliability of passive assessment of intervertebral motion in the cervical and lumbar spine: A systematic review. *Man Ther.* 2005 Abril; 10: 256-269.
57. Billis EV, Foster NE, Wright CC. Reproducibility and repeatability: errors of three groups of physiotherapists in locating spinal levels by palpation. *Man Ther.* 2003; 8(4): 223-232.
58. Van Dillen LR, Sahrmann SA, Norton BJ, Caldwell CA, Fleming DA, McDonnell MK, et al. Reliability of physical examination items used for classification of patients with low back pain. *Phys Ther.* 1998 Setembre; 78(9): 979-988.

59. Kilby J, Heneghan NR, Maybury M. Manual palpation of lumbo-pelvic landmarks: A validity study. *Man Ther.* 2001 En prensa; 1-4.
60. van Suijlekom HA, de Vet HCW, van der Berg SGM, Weber WEJ. Interobserver reliability in physical examination of the cervical spine in patients with headache. *Headache.* 2000 Juliol/Agost; 581-586.
61. Stovall BA, Shrawan K. Reliability of bony anatomic landmark asymmetry assessment in the lumbopelvic region: application to osteopathic medical education. *J Am Osteopath Assoc.* 2010 Nov; 110(11): 667-674.
62. Cook C, Hegedus E. Diagnostic utility of clinical test for spinal dysfunction. *Man Ther.* 2001 Febrer; 16(1): 21-25.
63. Robinson HS, Brox JI, Robinson R, Bjelland E, Solem S, Telje T. The reliability of selected motion- and pain provocation tests for the sacroiliac joint. *Man Ther.* 2007 Febrer; 12(1): 72-79.
64. Smedmark V, Wallin M, Arvidsson I. Inter-examiner reliability in assessing passive intervertebral motion of the cervical spine. *Man Ther.* 2000 Maig; 5(2): 97-101.
65. Fjellner A, Bexander C, Faleij R, Strender LE. Interexaminer reliability in physical examination of the cervical spine. *J Manipulative Physiol Ther.* 1999 Octubre; 22(8): 511-516.
66. Robinson R, Robinson HS, Bjørke G, Kvale A. Reliability and validity of a palpation technique for identifying the spinous processes of C7 and L5. *Man Ther.* 2009; 14: 409-414.
67. Johansson K, Ivarson S. Intra- and interexaminer reliability of four manual shoulder maneuvers used to identify subacromial pain. *Man Ther.* 2009; 14: 231-239.
68. Gonella C, Paris SV, Kutner M. Reliability in Evaluating Passive Intervertebral Motion. *Phys Ther.* 1982 Abril; 62: 436-444.
69. Maher CG, Latimer J, Adams R. An investigation of the reliability and validity of posteroanterior spinal stiffness judgments made using a reference-based protocol. *Phys Ther.* 1998 Agost; 78(8): 829-837.
70. Pownall PJ, Moran RW, Stewart AM. Consistency of standing and seated posture of asymptomatic male adults over a one-week interval: A digital camera analysis of multiple landmarks. *Int J Osteopath Med.* 2008 Febrer; 11: 43-51.

71. García García C. Validación del Test de Movilidad Mandibular Comparado con Tomografía Computerizada. Escuela de Osteopatía de Madrid..
72. McPartland JM, Goodridge JP. Counterstrain and traditional osteopathic examination of the cervical spine compared. *J Bodyw Mov Ther.* 1997 Abril; 1(3): 173-178.
73. Rey Eriz G, Albuquerque Sendín F, Barrera Mellado I, Martín Vallejo FJ. Validez del test de deslizamiento posteroanterior cervical medio respecto a análisis radiológico dinámico. *Osteopatía Científica.* 2009; 4(1): 13-19.
74. Paulet T, Fryer G. Inter-examiner reliability of palpation for tissue texture abnormality in the thoracic paraspinal region. *Int J Osteopath Med.* 2009 Juliol; 12: 92-96.
75. Potter L, McCarthy C, Oldham J. Intraexaminer reliability of identifying a dysfunctional segment in the thoracic and lumbar spine. *J Manipulative Physiol Ther.* 2006 Març/Abril; 29(3): 203-207.
76. Cook C, Stickley L, Akram N, Benavides Y, Renz C, Ramey K. Inter-Rater Reliability of Coupling Pattern Observations of the Pathological Lumbar Spine: A Pilot Study. *J Man Manip Ther.* 2004; 12(4): 192-198.
77. Degenhardt BF, Johnson JC, Snider KT, Snider EJ. Maintenance and improvement of interobserver reliability of osteopathic palpatory test over a 4-month period. *J Am Osteopath Assoc.* 2010 Octubre; 110(10): 579-586.
78. Snider KT, Snider EJ, Degenhart BF, Johnson JC, Kribs JW. Palpatory accuracy of lumbar spinous processes using multiple bony landmarks. *J Manipulative Physiol Ther.* 2011 Juny; 34(5): 306-313.
79. Bengaard K, Bogue RJ, Crow WT. Reliability of diagnosis of somatic dysfunction among osteopathic physicians and medical students. *Osteopathic Family Physician.* 2011; En premsa.
80. Kmita A, Lucas NP. Reliability of physical examination to assess asymmetry of anatomical landmarks indicative of pelvic somatic dysfunction in subjects with and without low back pain. *Int J Osteopath Med.* 2008 Gener; 11: 16-25.
81. O'Haire C, Gibbons P. Inter-examiner and intra-examiner agreement for assessing sacroiliac anatomical landmarks using palpation and observation: pilot study. *Man Ther.* 2000; 5(1): 13-20.
82. Vicent-Smith B, Gibbons P. Inter-examiner and intra-examiner reliability of the standing flexion test. *Man Ther.* 1999; 4(2): 87-93.

83. Bazzolo RJL, Legal L. Validación del test de movilidad en flexión-extensión de tobillo para la articulación tibioperonea superior comparado con radiología. *Osteopatía científica*. 2010; 5: 32-37.
84. Calderón C. Criterios de calidad en la investigación cualitativa en salud. *Rev Esp Salud Pública*. 2002 Setembre/Octubre; 76: 473-482.
85. Álvarez Solar M, López González ML, Cueto Espinar A. Indicadores bibliométricos, análisis temático y metodológico de la investigación publicada en España sobre epidemiología y salud pública (1988-1992). *Med Clin*. 1998; 111(14): 529-535.
86. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Towards Complete and Accurate Reporting of Studies of Diagnostic Accuracy: The STARD Initiative. *Clin Chem*. 2003 Gener; 49(1): 1-16.
87. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Clin Chem*. 2003 Gener; 49(1): 7-18.
88. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol*. 2003 Novembre; 3(1): 25.
89. Whiting PF, Weswood ME, Rutjes AW, Reitsma JB, Bossuyt PN, Kleijnen J. Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. *BMC Med Res Methodol*. 2006 Març; 6-9.
90. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. *Biometrics*. 1977 Març; 33(1): 159-174.
91. Landis RJ, Koch GG. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*. 1977 Juny; 33: 363-374.
92. Hopkins W. A new view of statistics. Internet Society for Sports Science. [Pàgina web]. [consulta: 30 Novembre 2011]. Disponible a: <http://www.sportsci.org/resource/stats/>.
93. García García C, Rodríguez Díaz J. Validación del test de movilidad mandibular comparado con tomografía computarizada. *Osteopatía científica*. 2008; 3(2): 58-64.
94. Stovall BA, Bae S, Kumar S. Anterior superior iliac spine asymmetry assessment on a novel pelvic model: an investigation of accuracy and

- reliability. *J Manipulative Physiol Ther.* 2010 Juny; 33(5): 378-385.
95. Arab AM, Abdollahi I, Joghataei MT, Golafshani Z, Kazemnejad A. Inter- and intra-examiner reliability of single and composites of selected motion palpation and pain provocation tests for sacroiliac joint. *Man Ther.* 2009 Abril; 14(2): 213-221.
 96. Berthelot JM, Labat JJ, Le Goff B, Gouin F, Maugars Y. Provocative sacroiliac joint maneuvers and sacroiliac joint block are. *Joint Bone Spine.* 2006 Gener; 73(1): 17-23.
 97. Haas M. The reliability of reliability (review). *J Manipulative Physiol Ther.* 1991; 14(3): 199-208.
 98. Seffinger M, Adams A, Najm W, Dickerson V, Mishra SI, Reinsch S, et al. Spinal palpatory diagnostic procedures utilized by practitioners of spinal manipulation: annotated bibliography of content validity and reliability studies. *J Can Chiropr Assoc.* 2003 Juny; 47(2): 93-109.
 99. Holsgaard-Larsen A, Myburgh C, Hartvigsen J, Rasmussen C, Hartvig M, Marstrand K, et al. Standardized simulated palpation training – Development of a Palpation Trainer and assessment of palpatory skills in experienced and inexperienced clinicians. *Man Ther.* 2010 Gener; 15: 254-260.
 100. Freburger JK, Riddle DL. Using Published Evidence to Guide the Examination of the Sacroiliac Joint Region. *Phys Ther.* 2001 Maig; 81(5): 1135-1143.
 101. Riddle DL, Freburger JK. Evaluation of the Presence of Sacroiliac Joint Region Dysfunction Using a Combination of Tests: A Multicenter Intertester Reliability Study. *Phys Ther.* 2002 Agost; 82(8): 772-781.
 102. Liebenson C, Lewit K. Palpation's reliability: a question of science vs. art? *J Bodyw Mov Ther.* 2003 Gener; 7(1): 46-48.
 103. Sarrado JJ, Clèries X, Ferrer M, Kronfly E. Evidencia científica en medicina: ¿única alternativa? *Gac Sanit.* 2004 Març; 18(3): 235-244.
 104. Korr IM. Osteopathic research: The needed paradigm shift. *JAOA.* 1991 Febrer; 91(2).
 105. Scientific European Federation of Osteopaths. Scientific European Federation of Osteopaths. [Pàgina web]. [consulta: 07 Novembre 2010]. Disponible a: www.scientific-european-federation-osteopaths.org/.
 106. Young S, Aprill C, Laslett M. Correlation of clinical examination characteristics with three sources of chronic low back pain. *Spine J.* 2003 Novembre-

Desembre; 3(6): 460-465.

107. Viikari-Juntura E, Takala EP, Riihimäki H, Martikainen R, Jäppinen P. Predictive validity of symptoms and signs in the neck and shoulders. *J Clin Epidemiol.* 2000 Agost; 53(8): 800-808.
108. Strimpakos N. The assessment of the cervical spine. Part 1: Range. *J Bodyw Mov Ther.* 2011; 15: 114-124.
109. Strimpakos N, Sakellari V, Gioftsos G, Kapreli E, Oldham J. Cervical joint position sense: an intra- and inter-examiner reliability study. *Gait Posture.* 2006; 23: 22-31.
110. Schneider M, Erhard R, Brach J, Tellin W, Imbarlina F, Delitto A. Spinal Palpation For Lumbar Segmental Mobility and Pain Provocation: An Interexaminer Reliability Study. *J Manipulative Physiol Ther.* 2008 Juliol/Agost; 31(6): 465-473.
111. Sancho R. Indicadores bibliométricos utilizados en la evaluación de la ciencia y la tecnología. Revisión bibliográfica. *Rev Esp Doc Cient.* 1990; 13(3-4): 842-865.
112. Patrick E, Fleiss S, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull.* 1979 Març; 86(2): 420-428.
113. Maher C, Adams R. Reliability of Pain and Stiffness Assessments in Clinical Manual Lumbar Spine Examination. *Phys Ther.* 1994 Setembre; 74(9): 801-809.
114. Lesho EP. An Overview of Osteopathic Medicine. *Arch Fam Med.* 1999 Desembre; 8: 477-484.
115. Johnston WL. Interexaminer reliability studies: Spanning a gap in medical research. *J Am Osteopath Assoc.* 1982 Agost; 81(12): 819-829.
116. Humpage C. Opinions on research and evidence based medicine within the UK osteopathic profession: A thematic analysis of public documents 2003e2009. *Int J Osteopath Med.* 2011 Juny; 14(2): 48-56.
117. Holmgren U, Waling K. Inter-examiner reliability of four static palpation tests used for assessing pelvic dysfunction. *Man Ther.* 2008; 13: 50-56.
118. Haneline M, Cooperstein R, Young M, Birkeland K. An annotated bibliography of spinal motion palpation reliability studies. *J Can Chiropr Assoc.* 2009 Març; 53(1): 40-58.

119. Guillory VJ. Research development at the University of Health Sciences College of Osteopathic Medicine. *J Am Osteopath Assoc.* 2001 May; 101(5): 265-267.
120. Cook C, Cleland J, Huijbregts P. Creation and Critique of Studies of Diagnostic Accuracy: Use of the STARD and QUADAS Methodological Quality Assessment Tools. *J Man Manip Ther.* 2007; 15(2): 93-102.
121. Horneij E, Hemborg B, Johnsson B, Ekdahl C. Clinical test on impairment level related to low back pain: a study of test reliability. *J Rehabil Med.* 2002; 34: 176-182.

Annexes

Taula 1: Adaptació del sistema QUADAS d'avaluació de la qualitat de la precisió dels articles.

	Si	No	Dubtós
1. Era l'espectre de pacients representatiu dels pacients que rebran el test (la prova estudiada) en la pràctica?			
2. Van ser els criteris de selecció clarament descrits?			
3. És l'estàndard de referència el millor per classificar correctament a les condicions de l'objectiu?			
4. És el període de temps entre l'estàndard de referència i el test suficientment curt, per ser raonablement segur el fet de que la condició de l'objectiu no canviï entre les dos proves?			
5. Ha estat verificada tota la mostra, o una selecció aleatòria de la mostra, utilitzant l'estàndard de referència de diagnòstic?			
6. Van rebre els pacients el mateix patró de referència, independentment del resultat de la prova?			
7. Era l'estàndard de referència independent del test (és a dir, el test avaluat no formava part de l'estàndard de referència)?			
8. Va ser l'execució del test descrita amb suficient detall per permetre la replicació de la prova?			
9. Va ser l'execució de la norma de referència descrita amb suficient detall per permetre la seva replicació?			
10. Van ser els resultats dels test interpretats sense el coneixement dels resultats del patró de referència?			
11. Van ser els resultats estàndard de referència interpretats sense el coneixement dels resultats del test?			
12. Eren les dades clíniques disponibles, a l'hora d'interpretar els resultats, les mateixes que estaran disponibles quan la prova sigui utilitzada a la pràctica?			
13. Van ser descrits els resultats no interpretables o entremitjos ?			
14. S'explicaquen les limitacions i punts febles de l'estudi?			

Taula 5: Quadre de valoració qualitativa segons criteri QUADAS

Regió	Autor principal i any	Ítem 1	Ítem 2	Ítem 3	Ítem 4	Ítem 5	Ítem 6	Ítem 7	Ítem 8	Ítem 9	Ítem 10	Ítem 11	Ítem 12	Ítem 13	Ítem 14
<i>Postura</i>	Pinal ⁷⁰ (2008)	N	S	?	S	S	S	N	S	S	?	N	S	S	S
<i>Osteopatia Cranial</i>	Moran ⁴⁷ (2001)	S	S	N	S	S	S	N	S	S	?	N	S	N	S
	Sergueef ⁴⁹ (2010)	N	S	N	S	S	S	N	S	N	?	N	N	S	S
	Somerfeld ⁴⁸ (2003)	S	S	?	S	S	S	N	S	S	?	N	S	S	S
<i>ATM</i>	García García ⁷¹ (2008)	S	S	S	?	S	S	S	S	S	N	N	S	?	N
<i>Columna cervical</i>	McPartland ⁷² (1997)	S	S	N	?	S	S	N	S	S	N	N	?	S	N
	Rey Eiriz ⁷³ (2009)	S	S	S	?	S	S	S	S	S	S	N	S	?	S
<i>Columna toràcica</i>	Ghoukassian ³⁴ (2001)	N	S	N	S	S	S	N	S	S	?	N	?	S	S
	Paulet ⁷⁴ (2009)	N	S	N	?	S	S	N	S	S	?	?	N	S	S
<i>Columna toràcica i lumbar</i>	Potter ⁷⁵ (2006)	N	S	N	S	S	S	N	S	S	?	S	S	S	N
<i>Columna lumbar</i>	Cook ⁷⁶ (2004)	N	S	N	?	S	S	S	?	N	N	N	N	S	N
	Degenhardt ³² (2005)	S	S	N	N	N	N	N	S	S	S	N	?	S	S
	Degenhardt ⁷⁷ (2010)	S	S	N	?	S	N	N	S	S	S	N	N	S	S
	Snider ⁷⁸ (2011)	S	S	S	S	S	S	N	S	S	N	S	?	S	S
	Spring ³³ (2001)	N	S	N	S	S	S	N	S	S	?	N	?	S	?
<i>Pelvis</i>	Bengaard ⁷⁹ (2011)	N	N	N	S	S	N	S	S	S	S	?	N	S	S
	Fryer ²⁶ (2005)	N	N	N	S	S	S	N	S	S	N	?	N	S	?
	Kmita ⁸⁰ (2008)	S	S	N	S	S	S	N	S	S	?	?	N	S	S
	O'Haire ⁸¹ (2000)	N	S	N	S	S	S	N	S	S	?	?	N	S	N
	Rajendran ³¹ (2011)	N	N	N	S	S	S	N	S	S	S	S	N	N	N
	Stovall ⁹⁴ (2010)	N	N	?	S	S	S	N	S	S	S	S	N	S	S
	Tong ⁴⁵ (2006)	N	N	N	S	S	S	N	S	S	N	N	N	S	S
	Vicent-Smith ⁸² (1999)	S	N	N	S	S	S	N	S	S	?	N	N	S	S
<i>Turmells</i>	Bazzolo ⁸³ (2010)	S	S	S	?	S	S	S	S	S	N	?	S	N	N
	Fryer ²⁷ (2006)	N	S	N	S	N	N	N	S	S	N	?	N	S	N
S=Si, N=No, ?=Dubtós		=													

Nota: Els ítems corresponen a les preguntes del sistema QUADAS de validació (Taula 1)

Taula 6: Característiques dels estudis inclosos distribuïts per regions i en ordre alfabètic del primer autor.

Regió	Autor principal i any	Participants	Examinadors	Procediment	Criteris i escales	Resultats	Conclusions dels autors
Postura	Pownall (70) (2008)	11 ♂, 29,6 ±10,4 anys, amb nivell de dolor o desconfort baix (escala McGill)	??	Enregistrament postura en bipedestació (anterior, posterior i lateral) i sedestació (lateral) de 29 variables en tres sessions espaiades en una setmana	Es marquen determinats relleus anatòmics amb adhesius i TiO ₂ i es mesuren angles i distàncies establerts i s'enregistren i es comparen els segons t=0, 20, 40, 60	17 variables ICC≥0.7 9 var 0.5≤ICC≤0.7 3 var 0.3≤ICC≤0.5	Les dades mostren alts nivells de fiabilitat per variables posturals en la vista anterior i posterior en bp, i per l'anteriorització del cap en bp i sd
Osteopatia Cranial	Moran (47) (2001)	11 voluntaris "sans" (5 ♂, 6 ♀) 25,8±8 anys	2 osteòpates postgraduats	Els examinadors palpen simultàniament el crani i el sacre i al moment de la fase de màxima flexió activen un pedal connectat amb un PC	Cada examinador palpa dos cops al crani o i al sacre abans d'intercanviar la posició. Es fan probes de 2 minuts i es realitzen a cec entre ells. Cadascú activa el mecanisme quan creu estar a la fase de màxima flexió	La mitjana de cicles MRP en un minut per l'examinador A es 2,92 al crani i 3,57 al sacre. El B 4,09 al crani i 4,17 al sacre Fiabilitat inter (CCI) de -0.09 a +0.31 Fiabilitat intra (CCI) A de +0.47a + 0.65 B de +0.52 a +0.73	La fiabilitat intraexaminador és de justa a bona, però la interexaminador de la palpació simultània és de pobre a no existent, amb lo que no es pot validar la hipòtesis del "core-link" en aquest estudi
	Sergueef (49) (2010)	734 estudiants graduats i postgraduats en osteopatia (es desconeix edat/sexe..)	734, els mateixos (uns es palpen als altres)	Es divideix el grup en dos meitats on uns fan d'observador i els altres de subjecte i s'intercanvien. Palpen durant tres minuts i contenen els cicles del MRP/min.	Cec a resultats de l'altre. La prova dura 3min en que han de contar el cicles MRP bifàsics i es divideix per tres per tenir cicles/min	Mitjana de 6.88±4.45 cicles/minut MRP Es divideixen segons experiència i els de un any d'experiència 7.39±4.70, els de 2 anys d'experiència 6.46±4.10. Els de 3 a 25 anys d'experiència 4.78±2.57	S'hauria de re considerar el valor de MRP de 8-14 cicles a 2-7 cicles. Com més experiència freqüència mes baixa i "estreta".
	Somerfeld (48) (2003)	49 subjectes sans 37.45±7.52 anys 39 ♂, 15♀	2 osteòpates postgraduats amb 300 hores d'entrenament i 7 anys d'experiència	Els examinadors palpen simultàniament el crani i el sacre separats amb una cortina (cec). Dos palpacions, una al crani i l'altre al sacre per subjecte.	Accionen un pedal quan creuen estar al inici de la fase de flexió i el deixen al inici de la fase d'extensió durant 90 segons.	Amb un 95% d'interval de confiança no es pot descriure el nivell de fiabilitat inter/intraexaminador. Es calculen 7 cicles/90 segons.	El MRP no es pot palpar de forma fiable i sembla estar influenciat per la respiració de l'examinador.

ATM articulació temporo mandibular	García García (93) (2008)	44 subjectes no aleatoritzats entre 18 i 49 anys, dividits en grup simptomàtic i asimptomàtic	3 osteòpates entrenats en diagnòstic d'ATM	Es realitza test actiu, test passiu, escalímetre (per mesurar obertura) i algòmetre. Es comparen resultats amb TC informàticament (escaneig imatges). TC es la prova de referència o "gold standard"	Es valoren les rotacions del còndil, desviacions de la trajectòria i sorolls i facilitat del moviment i si hi ha desviacions o sorolls es consideren positius de disfunció d'ATM: 9 categories de disfunció. Del TC s'escullen quatre imatges (obertura i tancament del còndil dret i esquerre) i es mesuren distàncies. Es comparen TA i TC.	Fiabilitat inter observador del test actiu: kappa 0.534 (0.41-0.6) establint concordança moderada Fiabilitat inter observador del test passiu: kappa 0.547 (0.41-0.6) establint concordança moderada Sensibilitat 97% Especificitat 72.7%	El test actiu de mobilitat mandibular és vàlid ja que es sensible a un 97% i específic en un 73% diagnosticant amb millor precisió als subjectes amb disfunció que sense. La fiabilitat inter observador és moderada.
Columna cervical	McPartland (72) (1997)	18 subjectes totals: 7 pacients de 39.1 anys de mitjana, 6♀ i 1♂ i 11 voluntaris asimptomàtics de 37.7 anys de mitjana, 6♀ i 5♂	2 osteòpates, un esquerrà de 10 anys d'experiència clínica i l'altre dretà de 40 anys d'experiència.	Es subjectes es palpen en decúbit supí amb ordre aleatoritzat d'examinadors. S'examinen els punts sensibles (tender points) descrits per Jones (9 parells). S'examinen després els paràmetres de la DS TART: canvis en els teixits, sensibilitat i restricció del moviment.	Els punts sensibles es palpen com petit nòdul o tensió miofascial i es marquen com presents o absents. En els cas de la diagnòstic TART es descriuen els tres paràmetres amb una magnitud del 0 al 10 on 0 es la no anormalitat i el 10 la màxima disfunció.	Fiabilitat inter-examinador: Punts sensibles pacients simptomàtics k=0.45 pacients asimptomàtics k=0.19 Diagnòstic TART pacients simptomàtics k=0.35 pacients asimptomàtics k=0.34	El diagnòstic segons punts sensibles és més fiable que el diagnòstic TART en pacients simptomàtics, però es menys fiable en pacients asimptomàtics.
	Rey Eiriz (73) (2009)	28 subjectes amb dolor mecànic d'un mes com a mínim 32.65±4.64 14 ♀, 14 ♂	1 osteòpata amb més de 5 anys d'experiència clínica	Es realitza un test de lliscament postero-anterior del raquis cervical inferior (TDPARCI) en sedestació i es compara amb radiografies dinàmiques cervicals en flexió i extensió dibuixant línees descrites per Maitland.	El TDPARCI es considera positiu si hi ha qualitat anormal del moviment, reproducció del dolor local o referit. (0 no hipomovilitat, 1 hipomovilitat). Le línees de les RX entre dos segments segueixen valors descrits per la mobilitat radiològica	Sensibilitat TDPARCI C3-C4 80% C4-C5 100% C5-C6 100% Valor predictiu negatiu C3-C4 93.75% C4-C5 100% C5-C6 100%	El TDPARCI és una prova diagnòstica vàlida per determinar pacients amb hipomovilitats cervicals i defineix que els pacients amb diagnòstic negatiu no presenten hipomovilitat.
Columna toràcica	Ghoukassian (34) (2001)	19 subjectes asimptomàtics ♂, 22 anys (19-40)	10 estudiants postgraduats d'osteopatia	Es marquen les espines dorsals amb un llapis alfabèticament de l'A a la L. Es fa una percussió sobre l'espina segment per segment.	Es consideren positius aquells on hi ha una alteració de la tensió dels teixits i que disminueix la recuperació dels teixits a la percussió	Fiabilitat interexaminador K=0.07, lleu	Els resultats demostren que la fiabilitat d'aquesta examinació de forma aïllada és qüestionable.

	Paulet (74) (2009)	10 subjectes amb símptomes toràcics, 19-23 anys, 5♂, 5 ♀	10 examinadors, 22-24 anys, 3♂, 7♀, estudiants d'últim any d'osteopatia. Entrenats una setmana abans durant una hora amb un osteopata amb experiència.	Realitzen una palpació longitudinal en un minut amb els dits índex i mig sobre quatre àrees predeterminades marcades amb retolador a la regió paravertebral toràcica, una almenys amb canvi de la textura .Les altres tres aleatòries.	Defineixen la disfunció somàtica amb les característiques de TART. La textura de l'àrea es "dura", "cordó" o "pantanós". Han d'escollir una, i sinó poden ser dos. Les anomenen d'un a 4 de cranial a caudal.	Fiabilitat interexaminador justa: k=0.26 Si es fa només sobre les cinc primeres palpacions per examinador continua sent justa: k=0.32	La fiabilitat de la palpació d'anormalitats a la regió paravertebral toràcica és només justa, fins i tot si es te en compte la suposada fatiga. L'avaluació dels canvis dels teixits és complexa.
Columna toràcica i lumbar	Potter (75) (2006)	12 subjectes estudiants universitaris, asimptomàtics i senes dolor, 35.7±11.95 anys 5 ♂, 7 ♀ . IMC: 23.9 Orde aleatoritzat	1 osteòpata amb 9 anys d'experiència clínica + 1 fisioterapeuta amb 14 anys d'experiència clínica.	L'osteòpata realitza un anàlisi visual postural, després mobilitat activa, després passiva per determinar àrees de hipo/hipermobilitat toràcica i lumbar. En decúbit pro es fa força anterior i cefàlica a les apòfisis espinoses. S'identifica. Es repeteix una hora més tard (sense activitats físiques)	Quan s'identifica es marca amb un bolígraf UV a la part inferior de l'espinosa superior. Es comprava amb un llum UV i es marca en un full d'acetat. Es fa dos cops (intentant cegar amb aleatorització) i es comparen acetats.	Columna toràcica CCI: 0.70 (0.27 a 0.90) Columna lumbar CCI: 0.96 (0.87 a 0.99)	Bona fiabilitat intraexaminador en la identificació de disfunció lumbar. La fiabilitat a la columna toràcica és de pobre a moderada.
Columna lumbar	Cook (76) (2004)	3 pacients P1 ♀ 21 anys, inestabilitat columna P2 ♂ 25 anys, hernia discal P3 ♀ 68 anys, degeneració disc	20 subjectes: 8 terapeutes manual (IAOM), 3 formació Maitland, 2 formació McKenzie, 1 osteopata, altres 6	Moviment acoblament de les vertebres lumbar de tres pacients en diferents posicions posturals en enregistraments on es marquen les vertebres L1-L2 i L4-L5..	Els observadors són cecs a la patologia i pacients i veuen els vídeos i en un qüestionari omplen els resultats observats sobre la si es fa primer rotació o lateroflexió del segment en diferents moviments.	Fiabilitat interexaminador Lateroflexió inicial: L1-L2: k=0.04 L4-L5: k=0.02 Rotació inicial: L1-L2: k=0.17 L4-L5: k=0.08	Fiabilitat inter examinador pobre en l'observació dels moviments acoblats de les vertebres lumbar, independent de la mostra d'observadors.
	Degenhardt (32) (2005)	119 subjectes dividits en: Grup 1: n=42, 26±4 anys, 31 ♂, 11♀ Grup 2: n=77, 31±8 anys, 30 ♂, 47♀	3 osteòpates amb entrenament de la matèria i menys de 10 anys d'experiència	Estudi de 8 diferents test sobre canvis dels teixits, sensibilitat i mobilitat i asimetria, distribuït en 3 fases: Fase 1: Fiabilitat interexaminador pre-entrenament. Fase 2: Entrenament per consensuar alguns tests. Fase 3: Fiabilitat interexaminador post-entrenament.	Es realitza un assaig a doble cec i els test que donen k menor de 0.40 es posen en comú durant sessions de una a dos hores a la setmana durant un període de 4 mesos fins arribar a acord. Es tornen a valorar els test de la fase dos en un altre assaig a doble cec.	Sensibilitat: Fase 1 k=0.34, justa Fase 3 k=0.68, substancial Canvis dels teixits: Fase 1 k=0.12, justa Fase 3 k=0.45, moderat Asimetria posicional: Fase 1 k=0.17, justa Fase 3 k=0.34, justa Pressió sobre espinoses: Fase 1 k=0.10, lleu Fase 3 k=0.20, lleu	Durant la fase inicial la fiabilitat inter observador total anava de -0.02 a 0.34. Després de l'entrenament de consens sobre test osteopàtics palpatoris comuns de la columna lumbar poden millorar la fiabilitat de tots fent que alguns siguin acceptables clínicament.
	Degenhardt (77) (2010)	64 subjectes simptomàtics (16) i asimptomàtics(48). 30 ±6 anys, 49♀, 15♂ N=33 sessió 1, 31 sessió 2, 17 sessió 5, 14 sessió 4, 17 sessió 5, 17 sessió 6.	2 osteòpates de l' AOA amb 10 i 3 anys d'experiència clínica. Es realitza entrenament previ.	S'avalua la regió lumbar (L1-L4) en decúbit pro amb quatre test: asimetria de les apòfisis transverses, canvis en els teixits, resistència a la pressió anterior i sensibilitat. Es realitzen 6 sessions a cec i es	Es defineixen tant la descripció dels test, com dels resultats que prèviament s'han consensuat. Es comparen els resultats. Si d'algun test falta fiabilitat es permet observar com el fan els dos però no es	Fiabilitat segons valor de k: Asimetria: 0.56 a 0.72, de moderada a substancial Canvis dels teixits: 0.23 a 0.55, moderada en 5 de les 6 sessions. Resistència a la pressió	Els examinadors són capaços de mantenir i millorar la fiabilitat interexaminador dels test palpatoris diagnòstics de la columna lumbar durant un període de 4 mesos.

			repeteixen els test 2 o 3 vegades i s'anota la presència o absència de troballes.	contabilitza.	anterior: 0.30 a 0.50 ,moderada en 3 de les 6 sessions Sensibilitat: 0.61 a 0.81 de substancial a perfecte en totes les sessions.		
	Snider (78) (2011)	60 subjectes (de 20 a 60 anys) 27♂, 33♀ IMC 25.6 kg/m2 (de 19.2 a 42.7, 6% obesos)	4 examinadors (3 osteòpates i 1 estudiant resident de 3er)	Tres examinadors independentment identifiquen el procés espinós de L1-L4 en decúbit pro fent servir diferents marques i relleus ossis. Es marquen a la pell amb marcadors i es fa RX AP.	Es determina un procediment per localitzar les diferents espinoses a partir de T12 i L5 i es marca amb un adhesiu i es fa RX. S'elimina i es torna a realitzar 3 vegades (una per cada participant).	Identificació correcta de la vertebra: k=0.81, CCI (0.79 a 0.83). Els osteòpates 67 a 78% correcte i l'estudiant 51%.	La localització de les espinoses lumbar amb l'ús de múltiples marques és més precís. La precisió depèn del nivell d'experiència, de la presència de anormalitats anatòmiques i de les característiques dels participants (l'obesitat).
	Spring (33) (2001)	10 ♀ asimptomàtiques, 21.5 anys	10 estudiants d'últim anys d'osteopatia amb una hora d'entrenament previ del mètode i dels resultats	Aleatoritzadament es palpen les apòfisis espinoses de L4 i després les apòfisis transverses en posició assentada, en decúbit pro i en postura d'esfinx. Es repeteix tres vegades amb espai de temps de 5 minuts.	Els examinadors escriuen les troballes com: Flexió dreta/extensió dreta/flexió esquerra/extensió esquerra/neutre esquerra	Fiabilitat intraexaminador: k= 0.04 (-0.14-0.16) nul·la a lleu. Fiabilitat interexaminador: k= 0.044, lleu	No hi ha acord en el diagnòstic posicional de la columna lumbar. La fiabilitat del diagnòstic posicional en tres postures de la columna lumbar és qüestionable.
Pelvis	Bengaard (79) (2011)	12 (2 subjectes per cada 25 examinadors, 6 canvis) i una pelvis de plàstic	151 estudiants d'osteopatia de diferents nivells	Cada examinador valorava els dos models i la pelvis de plàstic i anotava els resultats.	Es valora l'asimetria o no. Per definir disfunció l'asimetria havia ser de més d'un cm. Els resultats poden ser E superior, D superior o iguals	No hi ha valor k, ni CCI, ni r. El 89,2% dels examinadors escullen un costat a la pelvis fixa en que són iguals.	La palpació de l'asimetria no és una dada vàlida per definir DS.
	Fryer (26) (2005)	10 subjectes ♀ asimptomàtiques entre 22 i 24 anys	10 estudiants d'últim curs d'osteopatia (22 a 25a) 6♂,4♀. Un grup (n=5) no entrenat, l'altre entrenat 2 sessions d'1hora	Es valoren diferents punts anatòmics en 1 minut. Decúbit pro: EIPS, AIL (aspecte posterior e inferior) Decúbit supi: EIAS, MM TFA en sd. Es realitza tres cops de forma aleatoritzada.	Es determina si estan superior/inferior, anterior/posterior els punts o en el cas del TFA en sd si D=E, D>E,E>D (superior)	Fiabilitat intraexaminador: No entrenats, valor k MM= 0.59 (0.37-0.86) EIAS=0.19 (-0.01-0.40) EIPS=0.49 (0.22-0.76) TFA= 0.02 (-0.03-0.20) AILs/i=0.03 (-0.17-0.28) AILa/p=0.20 (-0.08-0.64) Entrenats MM= 0.68 (0.26-0.75) EIAS=0.65 (0.54-0.80) EIPS=0.54 (0.12-0.83) TFA= 0.41 (0.11-0.52) AILs/i=0.20 (-0.10-0.64) AILa/p=0.07 (-0.04-0.39)	Els examinadors que van fer la sessió d'entrenament tenen un augment dels valors de fiabilitat inter/intraexaminador, però la concordança continua sent baixa per acceptar-lo com a test clínic. No és un test vàlid i te una fiabilitat interexaminador pobre.

					<p>Fiabilitat interexaminador: No entrenats, valor k MM= 0.28; EIAS=-0.01; EIPS=0.15; TFA= 0.07; AILs/i=-0.01; AILa/p=-0.01 Entrenats MM= 0.31; EIAS=0.24; EIPS=0.08;TFA= 0.14 AILs/i=0.04; AILa/p=0.04</p>	
Kmita (80) (2008)	9 subjectes (21-78 anys) asimptomàtics i sense antecedents (n=4) i simptomàtics (n=5) 5 ♀, 4♂	4 examinadors (2 osteòpates i 2 estudiants d'últim curs) Entrenament d'una hora.	Simetria o asimetria de diferents referències anatòmiques: EIPS, SS, AIL, EIAS i MM de manera definida. Estudi cec. Es mesura 2 vegades cada marca de forma aleatoritzada.	Les asimetries de més de 3 mm es consideren positives i les de menys de 3 mm es consideren simètriques o negatives. Es valora la posició superior/inferior; o profund al SS.	<p>Fiabilitat intraexaminador k EIPS -0.29 a 0.39 SS: -0.28 a 0.83 AILp: -0.29 a 0.44 AILi: -0.29 a 0.34 EIAS: 0.25 a 0.63 MM: 0.20 a 1.00 Fiabilitat interexaminador k EIPS -0.38 a 0.35 SS: -0.34 a 0.26 AILp: -0.18 a 0.51 AILi: -0.03 a -0.1 EIAS: -0.13 a 0.5 MM: -0.05 a 0.49</p>	La fiabilitat interexaminador de l'examen físic és baixa, independentment de l'experiència. Les referències més fiables són l'EIAS i els MM. Els resultats no són consistents.
O'Haire (81) (2000)	10 ♀ asimptomàtiques de 18 a 30 anys	10 estudiants de 5é d'osteopatia amb un entrenament previ d'una hora	Es palpen EIPS, SS i AIL en DP 4 vegades de forma aleatoritzada i es registren en unes targetes.	Es valora la simetria o asimetria dels relleus: D>E, E>D, E=D	<p>Fiabilitat intraexaminador k: EPS: 0.33 (0.07 a 0.58) AIL: 0.21 (-0.05 a 0.69) SS: 0.24 (0.02 a 0.60) Fiabilitat interexaminador k: EPS: 0.04; AIL: 0.08; SS: 0.07</p>	La fiabilitat interexaminador és lleu i la intraexaminador és més alta però pobre. Potser s'explica per l'error en la localització de les referències.
Rajendran (31) (2011)	4 subjectes asimptomàtics 2 ♀ i 2♂	20 estudiants de 3er de la ESO, sense entrenament previ. Examen a doble cec i ordre aleatoritzat.	Examen a doble cec i ordre aleatoritzat. Es valoren en DP el SS, AIL, prova de flexió posterior i test del rebot lumbar	Es valora la profunditat i la qualitat segons uns barems	<p>Fiabilitat interexaminador k: SS: 0.033 AIL: -0.0476 PFP: -0.0425 RL: 0.0170</p>	Fiabilitat interexaminador pobre segons l'estudi basat en el model de diagnòstic de disfunció pèlvica de Mitchell.
Stovall (94) (2010)	Model pelvíc dissenyat expressament	10 examinadors: 5 estudiants de predoctorat en osteopatia i 5 osteòpates	Dos sessions d'avaluació de 13 "ajustaments" del model per valorar EIAS de forma aleatoritzada. Es valora individual i en grup. Es fan una sessions d'entrenament amb el model i es torna a valorar.	Es valora la superioritat de EIAS que es modifica aleatòriament segons 13 ajustaments	<p>Fiabilitat interexaminador abans d'entrenament individual k= 0.46, en grup k=0.37 Fiabilitat intraexaminador abans d'entrenament individual k=0.49 en grup k=0.52 Fiabilitat interexaminador després d'entrenament individual k=0.56</p>	Es desenvolupa un nou model pèlvic per poder consensuar la precisió i fiabilitat de l'avaluació de l'asimetria de les EIAS ja que es millora la precisió tant inter com intraexaminador.

					en grup k=0.53 Fiabilitat intraexaminador després d'entrenament individual k=0.56 en grup k=0.52		
	Tong (45) (2006)	24 subjectes simptomàtics 68.3±13.2 anys 15 ♀, 9 ♂	2 osteòpates	Es realitza exploració física per determinar disfunció sacre, posició del sacre i de la pelvis amb test sobre la SI i sobre relleus pèlvics segons palpació.	Es realitzen dos rondes i es recull si hi ha simetria o anormalitat del costat D o E. ES fan servir tres mètodes diagnòstics	Fiabilitat interexaminador Mètode 1: posició sacre k=0.47, posició pelvis k=0.08, costat disfunció sacre k=0.32 Mètode 2: posició sacre k=0.09, posició pelvis k=0.4, costat disfunció sacre k=0.16 Mètode 3: posició sacre k=0.16, posició pelvis k=0.1, costat disfunció sacre k=-0.33	Fent servir els resultats de les examinacions més fiables sempre es té la millor fiabilitat interexaminador.
	Vicent-Smith (82) (1999)	9 subjectes asimptomàtics de 21 anys de mitjana, 5 ♂, 4 ♀	9 estudiants postgraduats d'osteopatia amb experiència de 4-5 anys fent servir TFA	Realitzen el TFA i prèviament 10 moviments per minimitzar els falsos positius. Es realitza el TFA i s'enregistra el resultat. Es repeteix amb tots els subjectes aleatòriament.	Els test eren negatius si el moviment era simètric i positiu si un dels dos costats es movia més cefalícamment i i/o ventralment. Donen doncs: -, +D, +E.	Fiabilitat interexaminador: k= 0.052 (0.013 a 0.090) Fiabilitat intraexaminador: k= 0.46 (0.16 a 0.72)	El test de flexió anterior realitzat en pacients asimptomàtics demostra una fiabilitat interexaminador insignificant i moderada intraexaminador.
Turmells	Bazzolo (83) (2010)	51 subjectes entre 20 i 45 anys, simptomàtics 33 i asimptomàtics 18. 30 ♀, 21 ♂	3 osteòpates a doble cec	Es valora el test de flexo-extensió del turmell per l'articulació tibio-peronea superior. Es compara amb RX	Es valora la restricció del moviment o limitació durant la flexo-extensió. A les RX s'escanegen i es mesura la distància entre les línies que creuen meseta tibial i estiloides peroneal.	TMFET Sensibilitat 85% Especificitat 73.33% VPP: 69.9% VPN: 86.85 %	La prova és sensible i específica.
	Fryer (27) (2006)	5 subjectes ♀ asimptomàtiques de 22 a 24 anys per l'assaig 1 15 subjectes ♀ asimptomàtiques de 20 a 22 anys per l'assaig 2, però amb dismetria probable	5 estudiants d'últim curs de 22 a 24 anys, 2 ♂, 3 ♀. Es realitza entrenament previ de 20 min per consensuar	Els subjectes estirats aixequen la pelvis i la deixen caure, s'estiren les cames i es valora la dismetria. Al primer assaig no es pot consensuar si existeix o no, i es recluten subjectes amb dismetria per l'assaig 2. Es fa tres vegades	2 mm de dismetria es considera igual, més de 3 mm es considera no igual. Els resultats es recolliran com: D=E, D>E, E>D	Assaig 1: Fiabilitat interexaminador: k=0.22 Fiabilitat intraexaminador: k=0.31 (-0.20 a 0.73) Assaig 2: Fiabilitat interexaminador: k=0.94 Fiabilitat intraexaminador: k=0.94 (0.70 a 1.00)	La fiabilitat inter e intraexaminador és casi perfecte en subjectes seleccionats amb dismetria dels mal·lèols que suggereix que clínicament és fiable detectar asimetries dels mal·lèols d'uns 4 mm de diferència.

