


Universitat Autònoma de Barcelona

Facultat de Medicina

Departament de Pediatria, d'Obstetrícia i Ginecologia, i de
Medicina Preventiva



«Avaluació d'intervencions
terapèutiques no farmacològiques
en patologies del raquis. Alguns
reptes metodològics i del seu procés
d'avaluació, aprovació i aplicació al
sistema sanitari»

TESI DOCTORAL

Gerard Urrútia i Cuchí

Director: Dr. Xavier Bonfill i Cosp

Agost - 2011

Universitat Autònoma de Barcelona

Facultat de Medicina

Departament de Pediatria, d'Obstetrícia i Ginecologia, i de Medicina Preventiva

«Avaluació d'intervencions terapèutiques no farmacològiques en patologies del raquis. Alguns reptes metodològics i del seu procés d'avaluació, aprovació i aplicació al sistema sanitari»

TESI DOCTORAL

Gerard Urrútia i Cuchí

Director: Dr. Xavier Bonfill i Cosp

Agost - 2011

Universitat Autònoma de Barcelona

Facultat de Medicina

**Departament de Pediatria, d'Obstetrícia i Ginecologia, i de
Medicina Preventiva**

**«Avaluació d'intervencions
terapèutiques no farmacològiques
en patologies del raquis. Alguns
reptes metodològics i del seu procés
d'avaluació, aprovació i aplicació al
sistema sanitari»**

Gerard Urrútia i Cuchí

Agost de 2011

Memòria de Tesi com a compendi de publicacions presentada per Gerard Urrútia i Cuchí per optar al grau de Doctor en Medicina per la Universitat Autònoma de Barcelona i realitzada sota la direcció del Dr. Xavier Bonfill i Cosp.

Agraïments

A l'Esther, en Xavier, la Laia i la Jana, per les moltes hores d'absència -física o mental- robades del vostre benestar i que sempre heu sabut disculpar-me.

Als meus pares per haver sembrat en mi inquietuds i principis de vida d'un valor incalculable i etern.

Al Dr. Xavier Bonfill, amb qui porto treballant ininterrompudament des de fa vint anys, tota la meva carrera professional. La direcció d'aquesta tesi no és més que una molt petita part de tot el que, personalment i professional, he rebut generosament d'ell. Mai no estaré prou agraït!

A tots els meus companys i companyes del Servei d'Epidemiologia Clínica i Salut Pública – Centre Cochrane Iberoamericà de l'Hospital de la Santa Creu i Sant Pau, i abans de la Corporació Sanitària Parc Taulí de Sabadell, que m'han propiciat un ambient sempre idoni per a la realització professional i el creixement personal. Em sento molt afortunat de pertànyer a aquest gran equip!

Finalment, aquesta tesi deu també un agraïment al Dr. Francisco M. Kovacs. El que va començar com un encontre més o menys fortuït ha resultat en una fructífera relació professional que, de manera un tant inesperada, ha permès d'engendrar aquesta tesi.

«¹³ **V**aig aplicar-me de tot cor a cercar i a esbrinar amb saviesa tot el que s'esdevé en aquest món; tasca feixuga que Déu ha assenyalat als homes perquè se n'ocupin! ...
¹⁷ Vaig aplicar el meu cor a estudiar la saviesa i a comparar la niciesa i la follia; i he comprès que fins i tot això era afany inútil,
¹⁸ perquè on abunda la saviesa abunda el neguit, i qui augmenta la ciència augmenta el dolor. »

Paraules del rei Salomó al llibre d'Ecclésiastès cap. 1

ÍNDEX

1.- RESUM	5
2.- INTRODUCCIÓ	15
2.1.- Les patologies mecàniques del raquis.....	16
▪ La columna lumbar	
▪ Les patologies mecàniques del raquis	
▪ Importància social i sanitària de la lumbàlgia mecànica	
▪ El tractament de la lumbàlgia inespecífica	
▪ El tractament de l'estenosi espinal	
2.2.- L'avaluació dels tractaments i intervencions terapèutiques.....	33
▪ La medicina basada en l'evidència i els assaigs clínics	
▪ Les revisions sistemàtiques	
2.3.- Justificació de la tesi per compendi d'articles.....	38
3.- OBJECTIUS	41
3.1.- Objectius generals.....	42
3.2.- Objectius específics.....	42
4.- MÈTODES	45
5.- RESULTATS	51
5.1.- Publicacions presentades per aquesta tesi.....	52
5.2.- Resum dels resultats.....	53

- *Publicació nº 1*..... 53
 - *Urrútia G, Burton AK, Morral Fernández A, Bonfill Cosp X, Zanolli G. Neuroreflexotherapy for non-specific low-back pain. Cochrane Database of Systematic Reviews 2004, Issue 2. Art. No.: CD003009. DOI: 10.1002/14651858.CD003009.pub2.*
- *Publicació nº 2*..... 55
 - *Urrútia G, Kovacs FM, Nishishinya MB, and Olabe J. Percutaneous Thermocoagulation Intradiscal Techniques for Discogenic Low Back Pain. Spine (Phila Pa 1976). 2007;32:1146–54.*
- *Publicació nº 3*..... 57
 - *Kovacs FM, Urrútia G, Alarcón JD. Surgery versus conservative treatment for symptomatic lumbar spinal stenosis: a systematic review of randomized controlled trials. Spine (Phila Pa 1976). 2011;36:E1335-E1351.*
- *Publicació nº 4*..... 60
 - *Seco J, Kovacs FM, Urrútia G. The efficacy, safety, effectiveness, and cost/effectiveness of ultrasound and shock wave therapies for low back pain: A systematic review. Acceptat per a publicació a Spine J (2011).*
- *Publicació nº 5*..... 62
 - *Urrútia G, Bonfill X. Artículo especial. Análisis de la investigación española sobre dolencias del cuello y la espalda (1992–2006). Med Clin (Barc). 2010;135(5):215–21.*

6.- DISCUSSIÓ..... 65

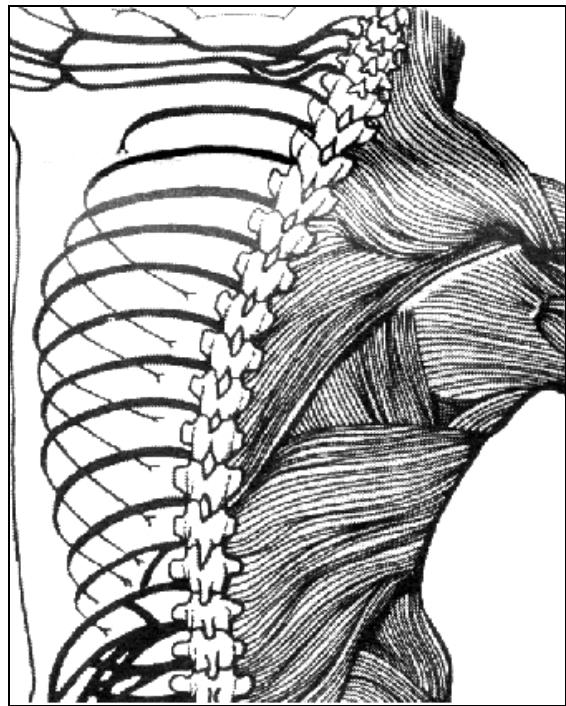
6.1.- Breu discussió específica derivada de les publicacions..... 66

- *Neurorreflexoteràpia per la lumbàlgia no específica*..... 66
- *IDET per la lumbàlgia discogènica*..... 68
- *Cirurgia per l'estenosi espinal lumbar simptomàtica*..... 70

▪ Ultrasons i ones de xoc per la lumbàlgia.....	73
▪ Investigació espanyola sobre patologies de l'esquena.....	74
6.2.- Discussió dels aspectes generals.....	76
▪ Utilitat, limitacions i reptes per dur a terme l'avaluació de les intervencions mitjançant revisions sistemàtiques.....	76
▪ Limitacions i reptes metodològics per dur a terme assaigs clínics que avaluen intervencions no farmacològiques.....	82
▪ Deficiències en el procés d'aprovació i introducció d'intervencions no farmacològiques per a les patologies de l'esquena al sistema de salut.....	89
7.- CONCLUSIONS.....	103
7.1.- Implicacions per a la pràctica.....	104
7.2.- Implicacions per a la recerca.....	104
8.- BIBLIOGRAFIA.....	107
9.- ANNEXOS.....	123
▪ <i>Annex 1: Urrútia G, Kovacs FM, Seco J. The efficacy, safety and effectiveness of laser therapy for low back pain: A systematic review. [Enviat a publicació a la revista Pain, al setembre de 2011].</i>	
▪ <i>Annex 2: Ferreira I, Urrútia G, Alonso-Coello P. Systematic Reviews and Meta-Analysis: Scientific Rationale and Interpretation. Rev Esp Cardiol. 2011;64(8):688–96.</i>	
▪ <i>Annex 3: Urrútia G, Bonfill X. Revisiones sistemáticas. En: J. Jiménez Villa, J.M. Argimon Pallàs, A. Martín Zurro, M. Vilardell Tarrés. Publicación científica biomédica. Cómo escribir y publicar un artículo de investigación. Barcelona: Elsevier; 2010. p. 229-45.</i>	

- **Annex 4:** *Urrútia G, Bonfill X. Declaración PRISMA: una propuesta para mejorar la publicación de revisiones sistemáticas y metanálisis. Med Clin (Barc). 2010 Oct 9;135(11):507-11.*
- **Annex 5:** *Carragee EJ, Deyo RA, Kovacs FM, Peul WC, Lurie JD, Urrútia G, Corbin TP, Schoene ML. Clinical research: is the spine field a mine field? Spine (Phila Pa 1976). 2009 Mar 1;34(5):423-30.*

1. Resum



1.- RESUM

Antecedents: Les patologies mecàniques del raquis són un problema de salut molt freqüent a la població general i treballadora, amb un elevat impacte social i sanitari. S'estima que la seva prevalença al llarg de tota la vida pot arribar a ser del 80%, aproximadament un 10% dels casos es cronifiquen i el seu cost equival al 2% del PIB. S'han proposat una gran quantitat d'intervencions no farmacològiques (conservadores, invasives i quirúrgiques) per al tractament d'aquestes patologies. Malgrat això, sovint els resultats no són satisfactoris. La base d'evidència científica de moltes d'aquestes intervencions és molt feble o desconeguda.

Mètodes: S'han portat a terme sengles revisions sistemàtiques per avaluar l'eficàcia i seguretat de quatre intervencions no farmacològiques sobre les quals hi havia alguna incertesa: 1) neuroreflexoteràpia (NRT) per al tractament de la lumbàlgia no específica en pacients adults, 2) tècniques d'electrotermocoagulació intradiscal percutànies (IDET i PIRFT) per al tractament de la lumbàlgia d'origen discogènic, 3) cirurgia en comparació amb el tractament conservador per al tractament de l'estenosi lumbar simptomàtica i 4) teràpies basades en ultrasons i ones de xoc per al tractament de la lumbàlgia.

També s'ha portat a terme un estudi descriptiu sobre la recerca a l'àmbit de les patologies mecàniques del raquis portada a terme a l'estat espanyol.

Resultats: 1) La NRT va millorar significativament el dolor, la discapacitat i la funció en comparació amb el placebo i també el tractament habitual. No hi van haver problemes de seguretat. L'efecte sembla mantenir-se a llarg termini. 2) Els resultats dels estudis sobre IDET són contradictoris i, per tant, els seus beneficis incerts. La seguretat del procediment no ha estat mesurada de forma rigorosa. 3) La cirurgia de descompressió va millorar significativament el dolor, la discapacitat i la funció en comparació amb el tractament conservador. L'efecte sembla mantenir-se a llarg termini. 4) Els ultrasons i les ones de xoc no varen ser millors que el placebo.

El nombre d'estudis sobre patologies mecàniques del raquis finançats per entitats públiques a Espanya és baix. Els criteris de selecció semblen inadequats i poc transparents.

Discussió: Les revisions sistemàtiques es demostren com a instruments útils per apropar l'evidència disponible a la presa de decisions clíniques i sanitàries. Les seves limitacions, sovint, venen imposades per problemes comuns detectats en la realització d'assaigs clínics sobre intervencions no farmacològiques així com la qualitat deficient de les seves publicacions.

La qualitat de les evidències disponibles en aquestes revisions és molt variable, però en el seu conjunt es pot qualificar de moderada o baixa, per les nombroses limitacions detectades a molts dels estudis originals així com també a les seves respectives publicacions. No obstant això, ha estat possible arribar a conclusions suficientment clares sobre l'efectivitat en tres d'elles (dues en sentit favorable i una en sentit negatiu) mentre que l'evidència roman encara incerta en la darrera.

El procés d'avaluació i control de les tecnologies i procediments introduïts al sistema de salut pel tractament de les patologies d'esquena és encara deficient, promovent-se l'ús d'intervencions amb poc, o cap, sòlid fonament científic, en detriment de la salut dels pacients i/o la sostenibilitat del sistema.

Conclusions: 1) Són nombrosos els tractaments i intervencions no farmacològiques que actualment es proposen als pacients amb patologies mecàniques d'esquena, moltes de les quals no estan suficientment avalades per estudis científics de qualitat que hagin demostrat la seva eficàcia, efectivitat i seguretat. 2) De les intervencions avalades a la tesi, dues han demostrat ser efectives i es poden recomanar: la NRT per al tractament de la lumbàlgia crònica i subaguda no específica, i la cirurgia de descompressió en determinats pacients de risc amb estenosi espinal simptomàtica (sobretot en presència de claudicació neurògena) que han fracassat a un tractament conservador previ després de 3 a 6 mesos. 3) L'evidència sobre l'efectivitat i seguretat de les tècniques percutànies d'electrotermocoagulació intradiscal (IDET) per al tractament de la lumbàlgia discogènica és encara no concloent i, per tant, no es poden recomanar. 4) Els ultrasons o les ones de xoc no semblen tenir cap

efecte favorable clínicament rellevant per al tractament de la lumbàlgia. 5) Cal seguir investigant en alguns aspectes relacionats amb les intervencions avaluades per les revisions com, per exemple: i) el tipus específic de cirurgia més efectiva (especialment el paper de la fusió instrumentada) i els factors pronòstics dels pacients amb estenosi espinal (refinament dels criteris d'indicació), ii) seguretat i confirmació dels resultats a llarg termini de la cirurgia, iii) confirmació dels resultats del dispositiu interespinal de tracció amb nous estudis (replicació), 4) cal més investigació sobre procediments i tècniques mínimament invasives, com l'IDET. 5) Les agències de finançament públic de la recerca haurien d'incrementar la seva inversió en l'àrea de les patologies de l'esquena i també revisar i millorar els seus mecanismes d'avaluació i selecció dels projectes, que haurien de ser més transparents. 6) Cal promoure la realització d'assaigs clínics d'intervencions no farmacològiques en l'àrea de les patologies de l'esquena (cirurgia, fisioteràpia, teràpia física, intervencions invasives, educatives o psicològiques, etc.) creant una cultura de l'avaluació i facilitant els mitjans per fer-ho. Així mateix, cal millorar la qualitat dels estudis i assaigs clínics en aquesta àrea, superant les nombroses mancances detectades, així com també la de les publicacions derivades, idealment en revistes indexades, seguint les recomanacions de la declaració CONSORT.

1.- RESUMEN

Antecedentes: Las patologías mecánicas del raquis son un problema de salud muy frecuente en la población general y trabajadora, con un elevado impacto social y sanitario. Se estima que su prevalencia a lo largo de toda la vida puede llegar a ser del 80%, aproximadamente un 10% de los casos se cronifican y su coste equivale al 2% del PIB. Se han propuesto una gran cantidad de intervenciones no farmacológicas (conservadoras, invasivas y quirúrgicas) para el tratamiento de estas patologías. Sin embargo, a menudo los resultados no son satisfactorios. La base científica de muchas de estas intervenciones es muy débil o desconocida.

Métodos: Se han llevado a cabo sendas revisiones sistemáticas para evaluar la eficacia y seguridad de cuatro intervenciones no farmacológicas sobre las que había alguna incertidumbre: 1) neuroreflexoterapia (NRT) para el tratamiento de la lumbalgia no específica en pacientes adultos, 2) técnicas de electrotermocoagulación intradiscal percutáneas (IDET y PIRFT) para el tratamiento de la lumbalgia de origen discogénico, 3) cirugía en comparación con el tratamiento conservador para el tratamiento de la estenosis lumbar sintomática y 4) terapias basadas en ultrasonidos y ondas de choque para el tratamiento de la lumbalgia.

También se ha llevado a cabo un estudio descriptivo sobre la investigación en el ámbito de las patologías mecánicas del raquis llevada a cabo en España.

Resultados: 1) La NRT mejoró significativamente el dolor, la discapacidad y la función en comparación con el placebo y también el tratamiento habitual. No hubo problemas de seguridad. El efecto parece mantenerse a largo plazo. 2) Los resultados de los estudios sobre IDET son contradictorios y, por tanto, sus beneficios inciertos. La seguridad del procedimiento no ha sido medida de forma rigurosa. 3) La cirugía de descompresión mejoró significativamente el dolor, la discapacidad y la función en comparación con el tratamiento conservador. El efecto parece mantenerse a largo plazo. 4) Los ultrasonidos y las ondas de choque no fueron mejores que el placebo.

El número de estudios sobre patologías mecánicas del raquis financiados por entidades públicas en España es bajo. Los criterios de selección parecen inadecuados y poco transparentes.

Discusión: Las revisiones sistemáticas se demuestran como instrumentos útiles para acercar la evidencia disponible en la toma de decisiones clínicas y sanitarias. Sus limitaciones, a menudo, vienen impuestas por problemas comunes detectados en la realización de ensayos clínicos sobre intervenciones no farmacológicas así como por la calidad deficiente de sus publicaciones.

La calidad de las evidencias disponibles en estas revisiones es muy variable, pero en su conjunto se puede calificar de moderada o baja, por las numerosas limitaciones detectadas en muchos de los estudios originales así como en sus respectivas publicaciones. Sin embargo, ha sido posible llegar a conclusiones suficientemente claras sobre la efectividad en tres de ellas (dos en sentido favorable y una en sentido negativo) mientras que la evidencia permanece aún incierta en la última.

El proceso de evaluación y control de las tecnologías y procedimientos introducidos al sistema de salud para el tratamiento de las patologías de espalda es aún deficiente, promoviendo el uso de intervenciones con poco, o ningún, sólido fundamento científico, en detrimento de la salud de los pacientes y / o la sostenibilidad del sistema.

Conclusiones: 1) Son muchos los tratamientos e intervenciones no farmacológicas que actualmente se proponen a los pacientes con patologías de espalda, muchas de las cuales no están suficientemente avaladas por estudios científicos de calidad que hayan demostrado su eficacia, efectividad y seguridad. 2) De las intervenciones evaluadas en la tesis, dos han demostrado ser efectivas y podrían recomendarse: la NRT para el tratamiento de la lumbalgia crónica y subaguda no específica, y la cirugía de descompresión en determinados pacientes de riesgo con estenosis espinal sintomática (sobre todo en presencia de claudicación neurógena) y que no han respondido a un tratamiento conservador por 3 a 6 meses. 3) La evidencia sobre la efectividad y seguridad de las técnicas percutáneas de electrocoagulación intradiscal (IDET) para el tratamiento de la lumbalgia de origen discogénico es no concluyente y,

por tanto, no se pueden recomendar. 4) Los ultrasonidos y las ondas de choque no parecen tener ningún efecto favorable clínicamente relevante para el tratamiento de la lumbalgia. 5) Hay que seguir promoviendo la investigación sobre algunos aspectos relacionados con las intervenciones evaluadas por las revisiones como, por ejemplo: i) el tipo específico de cirugía más efectiva (especialmente el papel de la fusión instrumentada) y los factores pronósticos de los pacientes con estenosis espinal (refinamiento de los criterios de indicación), ii) seguridad y confirmación de los resultados a largo plazo de la cirugía, iii) confirmación de los resultados del dispositivo de tracción interespinoso con nuevos estudios (replicación), iv) hay que fomentar más investigación sobre procedimientos y técnicas mínimamente invasivas, como l'IDET; 6) Las agencias públicas de financiación de la investigación deberían incrementar su inversión en e área de las patologías de la espalda y también revisar y mejorar sus mecanismos de evaluación y selección de los proyectos, que deberían ser más transparentes. 7) Hay que promover la realización de ensayos clínicos de intervenciones no farmacológicas en el área de las patologías de la espalda (cirugía, fisioterapia, terapia física, intervenciones invasivas, educativas o psicológicas, etc.) creando una cultura de la evaluación y facilitando los medios para ello. Asimismo, hay que mejorar la calidad de los estudios y ensayos clínicos en esta área, superando las numerosas limitaciones detectadas, así como también la de las publicaciones derivadas, idealmente en revistas indexadas, siguiendo las recomendaciones de la declaración CONSORT.

1.- ABSTRACT

Background: The mechanical disorders of the spine are a very common health problem in the general population and among workers with a high social and health impact. It is estimated that the prevalence throughout life may be 80%, approximately 10% of cases become chronic, and its cost is equivalent to 2% of the GDP. Numerous non-pharmacological interventions (conservative, surgical, and invasive) have been proposed for the treatment of these pathologies. However, often the results are not satisfactory. The scientific ground for many of these interventions is weak or unknown.

Methods: We conducted four systematic reviews to evaluate the efficacy and safety of four non-pharmacological interventions for which there was some uncertainty: 1) neuroreflexotherapy (NRT) for the treatment of nonspecific low back pain in adult patients, 2) percutaneous intradiscal electrothermocoagulation techniques (IDET and PIRFT) for the treatment of discogenic low back pain, 3) surgery compared with conservative treatment for symptomatic spinal stenosis, and 4) ultrasound based therapies for the treatment of low back pain.

We have also carried out a descriptive study on research on mechanical disorders of the spine carried out in Spain.

Results: 1) NRT significantly improved pain, level of disability, and function compared to placebo and standard care. There were not any concerns about safety. The effect seems to stay long term (observational follow-up). 2) The results of studies on IDET are at best conflicting and, therefore, its benefits remain uncertain. The safety of the procedure has not been measured appropriately. 3) Decompression surgery improved pain, level of disability, and function compared to conservative treatment. The effect seems to stay long term. 4) Ultrasound and shock waves were no better than placebo.

The number of studies on mechanical disorders of the spine funded by public bodies in Spain is low. The selection criteria seem to be inadequate and opaque.

Discussion: Systematic reviews prove to be a useful tool to bring evidence available to clinical and health care decision making. Their limitations often are imposed by common problems found in clinical trials on non-pharmacological interventions as well as by the poor quality of their publications.

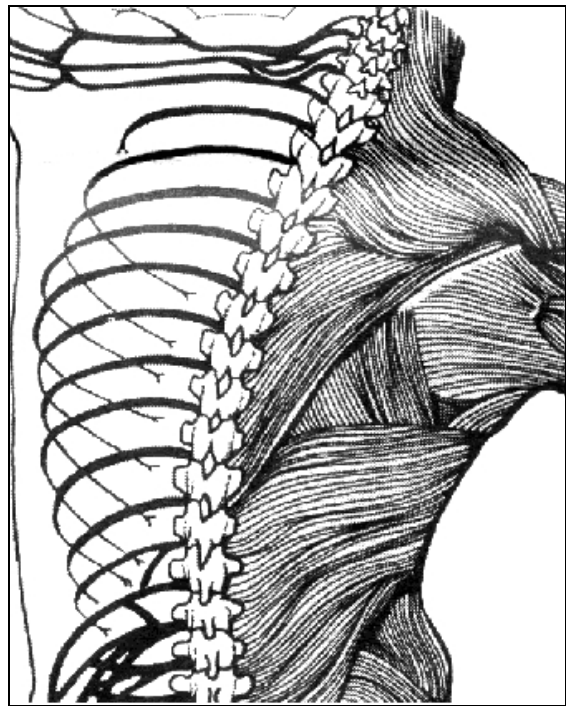
The quality of evidence available on the reviews of this thesis is highly variable, but as a whole can be described as moderate or low, due to the many constraints identified in many of the original studies as well as their respective publications. However, it has been possible to arrive at clear conclusions about the effectiveness of three of them (two in the positive direction and one in the negative) while the evidence remains uncertain in the last one.

The assessment and monitoring process for technologies and procedures for the treatment of back disorders introduced in the health system is still flawed. It promotes the use of interventions with little or no solid scientific basis, to the detriment of patient health and/or sustainability of the system.

Conclusions: 1) There are many non-pharmacological interventions currently being proposed to patients with back disorders, many of which are not sufficiently supported by sound scientific studies that demonstrate their efficacy, effectiveness, and safety. 2) Of the interventions evaluated in this thesis, two have proven to be effective and can be recommended: NRT for the treatment of chronic and subacute non-specific low back pain, and decompression surgery in selected patients at risk with symptomatic spinal stenosis (mainly with neurogenic claudication) that have not responded to conservative treatment for 3 to 6 months. 3) Evidence on the effectiveness and safety of percutaneous electrothermal intradiscal techniques (IDET) for the treatment of discogenic low back pain is not conclusive and therefore cannot be recommended. 4) Ultrasounds do not seem to have any clinically relevant effect for the treatment of low back pain. 5) It is necessary to promote research in general in this area, and in particular address some controversial aspects related to the interventions evaluated in this thesis: i) identification of the most effective type of surgery (specially the role of instrumented fusion) and prognostic factors in patients with spinal stenosis (refinement of the indication criteria for a better patient selection), ii) safety and long-term results, iii) confirmation of the preliminary positive

results of the interspinous traction device, and iv) promotion of sound research on minimally invasive procedures such as IDET. 6) Public funding bodies should increase their commitment to support research on back disorders. They should also review and improve the evaluation and selection process of research proposals that will be funded, which should be more transparent. 7) More high-quality studies (preferably clinical trials) assessing non-pharmacological interventions for low back pain (surgery, physiotherapy, physical therapy, invasive, educational or psychological interventions, etc.) must be encouraged, fostering a culture of evaluation and facilitating the conditions to do so. Likewise, the quality of clinical trials in this area should be increased, overcoming the numerous limitations identified in these reviews. The formal quality of the publications derived from these studies must also abide to CONSORT guidelines.

2. Introducció



2.- INTRODUCCIÓ

2.1.- Les patologies mecàniques del raquis

La columna lumbar

L'esquena és una estructura complexa formada per óssos, músculs i altres teixits que formen la part posterior del tronc, des del coll fins a la pelvis. La part central de l'esquena és la columna vertebral, que no només sosté el pes de la part superior del cos sinó que allotja i protegeix la medul·la espinal, la delicada estructura del sistema nerviós que transporta els senyals que controlen els moviments del cos i li transmeten les seves sensacions sensibles. La columna vertebral està formada per un conjunt de fins a 33 óssos (les vèrtebres) apilats l'un sobre l'altre. Cadascun d'aquests óssos conté un forat arrodonit que, alineat amb el de la resta de les vèrtebres, forma un canal que envolta a la medul·la espinal, la qual descendeix des de la base del cervell i s'estén en l'adult fins just per sota del tòrax. Nervis petits (arrels nervioses) entren i surten de la medul·la espinal a través d'espais entre les vèrtebres. Degut a que els óssos de la columna vertebral continuen creixent molt després de que la medul·la espinal assoleixi la seva longitud completa en la primera infància, les arrels nervioses de la zona lumbar i de les extremitats inferiors s'estenen molts centímetres cap avall per la columna vertebral abans de sortir. Aquest gran feix d'arrels nervioses es coneix amb el nom de cauda equina o cua de cavall. Els espais laterals que hi ha entre les vèrtebres, i per on entren i surten les arrels nervioses (foramen intervertebral o forat de conjunció), es mantenen per uns coixins de cartílag esponjosos i rodons anomenats discos intervertebrals que permeten la flexibilitat de l'esquena lumbar i actuen com a esmorteïdors en tota la columna vertebral per protegir els óssos que la formen quan es mou el cos. Els discos representen fins al 25% de l'alçada total de la columna vertebral. Bandes de teixit tou i elàstic (l·ligaments i tendons) mantenen a les vèrtebres al seu lloc i uneixen els músculs a la columna vertebral.

Començant per la part superior, la columna té quatre regions: i) les set vèrtebres cervicals o del coll (C1-C7), ii) les 12 vèrtebres toràciques o superiors (T1-T12), iii) les cinc vèrtebres lumbars (L1-L5), i iv) el sacre i el còccix, un grup

de nou óssos fusionats a la base de la columna. La **columna lumbar** suporta el pes corporal de la part superior del cos. Per aquest motiu, les cinc vèrtebres lumbar tenen una estructura molt robusta, tant més voluminoses quan més avall es troben situades. A més, són el darrer nivell mòbil de la columna. És per aquesta raó que la columna lumbar és la zona de l'esquena on amb més freqüència se sent dolor. Així, l'articulació lumbosacra, que suporta el major pes i la major força cisallant en flexió o extensió, és com una tija que es mou cap endavant, enrere, lateralment i en rotació sobre el punt de suport lumbosacre. L'últim espai lumbosacre (L5-S1) és el que té una major mobilitat (75% de tota la flexió de la columna lumbar, front a només el 5-10% entre L1 i L4). Això determina que la major exigència i la major quantitat de malaltia lumbar es generi a nivell de L5-S1 i L4-L5.

El dolor lumbar o **lumbàlgia** es defineix com un dolor o sensació de malestar que es localitza entre la darrera costella i la zona glútia, i que pot anar acompanyat o no d'irradiació a nivell de l'extremitat inferior.^(1,2)

Les patologies mecàniques del raquis

El dolor d'esquena pot ser degut a molt diverses causes, que poden tenir el seu origen a les estructures de l'esquena o bé externes. Cal tenir en compte que el dolor d'esquena no és un diagnòstic ni una malaltia, sinó que es tracta d'un símptoma i que, per tant, pot ser degut a múltiples alteracions de molt diversa gravetat i repercussió. En ocasions, el dolor lumbar pot indicar un problema mèdic important. Així, per exemple, el dolor acompanyat de febre, la pèrdua de control vesical o intestinal, o la debilitat progressiva a les cames poden indicar la presència d'una patologia greu (són els anomenats signes d'alarma o *red flags*). En aquests casos, quan el metge responsable sospita l'existència d'una causa específica per al dolor d'esquena, ha de prescriure les proves diagnòstiques necessàries per confirmar-la o descartar-la.

Taula 1. Causes de dolor lumbar.

Estructurals <ul style="list-style-type: none">• Mecànic o no específic• Osteoartritis de les superfícies articulars• Prolapse del disc intervertebral• Fissura anular del disc intervertebral• Espondilòlisi o espondilolistesi• Estenosi espinal	Infecció <ul style="list-style-type: none">• Discitis• Osteomielitis• Abscés paraespinal
Neoplàsia <ul style="list-style-type: none">• Primària o secundària	Inflamatòria <ul style="list-style-type: none">• Espondiloartropaties• Sacroiliïtis o disfunció sacroilíaca
Dolor referit a la columna <ul style="list-style-type: none">• Origen en una víscera, estructures retroperitoneals, sistema urogenital, aorta, o maluc	Metabòlica <ul style="list-style-type: none">• Col·lapse vertebral osteoporòtic• Malaltia de Paget• Osteomalàcia• Hiperparatiroidisme

Les **patologies (o síndromes) mecàniques del raquis (PMR)** constitueixen un conjunt de causes o afeccions pròpies de l'esquena (intrínseques) que poden produir dolor. Típicament són aquelles causes en les que el dolor d'esquena varia en funció de la postura i tendeix a empitjorar amb l'esforç físic. Són la causa més freqüent de dolor d'esquena amb gran diferència, en comparació amb les causes extrínseques que tenen el seu origen en estructures alienes a la columna. Inclouen diagnòstics com l'artrosi, l'hèrnia discal, la protrusió discal, la contractura muscular, l'escoliosi, l'espondilolistesi o l'estenosi espinal, entre d'altres.

Les PMR es defineixen com la presència de dolor en la zona vertebral o paravertebral, localitzat entre la base del crani i el límit inferior de les natges, habitualment amb limitació dolorosa de la mobilitat de la columna i acompanyat o no de dolor referit o irradiat. La intensitat del dolor es veu influenciada per la postura, els moviments i l'esforç (tendeix a empitjorar amb l'esforç i a millorar amb el repòs dels segments implicats), sense que hi hagi dolors nocturns que despertin el pacient. El seu diagnòstic requereix haver descartat que el dolor es degui a fractures, traumatismes directes, espondilitis (infecció i inflamació de

les articulacions vertebrals), malalties sistèmiques (com tumors, infeccions o afeccions vasculars), afeccions psicosomàtiques i a dolor irradiats d'origen extern a la columna vertebral i la seva musculatura (com dolors referits d'origen digestiu, pulmonar o ginecològic).^(3,4)

Generalment, la **lumbàlgia mecànica** és atribuïda a alteracions estructurals de la columna (per exemple, canvis degeneratius relacionats amb l'edat, hèrnia o protrusió del disc, estenosi espinal o espondilolistesi), a una sobrecàrrega funcional (per exemple, aixecament repetit de pesos, vibracions excessives o conducció prolongada de vehicles) o postural (per exemple, postures mantingudes inadequades de sedestació o per dur a terme activitats quotidianes sense tenir en compte les normes bàsiques d'ergonomia). No és clar que l'estil de vida sedentari pugui causar, per si mateix, lumbàlgia mecànica.⁽⁵⁾

S'estima que fins al 85% de les lumbàlgies mecàniques són de **tipus inespecífic** (dolor que no pot ser atribuït a una causa específica identificable), mentre que aproximadament el 15% presenten una patologia identificable (principalment hèrnia de disc lumbar i estenosi espinal -amb o sense espondilolistesi-).^(3,4)

La fisiopatologia de la lumbàlgia mecànica és complexa i multifacètica. Se sospita que nombroses estructures anatòmiques de la columna lumbar (óssos, lligaments, tendons, discs, músculs) hi estan implicades. Molts d'aquests components tenen una innervació sensorial que pot generar senyals nociceptives en resposta a estímuls d'agressió als teixits. Altres causes poden ser neuropàtiques (per exemple, la ciàtica). La majoria de casos de lumbàlgia crònica molt probablement impliquen una etiologia mixta nociceptiva i neuropàtica. Addicionalment, estudis portats a terme a finals del segle XX suggereixen que hi hauria causes químiques que també hi juguen un paper en la producció de lumbàlgia mecànica (per exemple, components del nucli polpós del disc intervertebral com l'enzim fosfolipasa A2, el glutamat o la substància P).

També s'ha proposat la implicació de factors psicosocials en l'aparició i sobretot en la progressió de la lumbàlgia. Nombrosos autors han posat èmfasi en el paper del malestar psicològic manifestat com ansietat, depressió, catastrofització o somatització, com a factor de risc de lumbàlgia. També factors relacionats amb el treball i la professió, tant físics (per exemple, carregar pesos, bipedestació prolongada, vibració corporal, etc.) com psíquics (per exemple, treball estressant o monòton, insatisfacció laboral, etc.), han estat associats o suggerits com a factors predictius de lumbàlgia i de la seva cronificació.⁽⁶⁻¹⁴⁾ En conseqüència, l'estrès social i laboral i la insatisfacció amb el treball són factors que cal estudiar de forma individualitzada en cada treballador amb lumbàlgia. Aquests factors també poden modificar les respostes al tractament i el símptomes una vegada instaurada la síndrome.

Importància sanitària i social de la lumbàlgia mecànica

El dolor d'esquena constitueix una experiència gairebé universal. Les estimacions de la seva prevalença varien molt entre estudis, probablement degut a diferències de mostreig, metodològiques i, en part també, a variacions geogràfiques. La prevalença instantània de lumbàlgia mecànica estimada en alguns països europeus arriba fins al 33% a Bèlgica (mentre que a Suècia és del 12%, a Dinamarca del 13,7% i al Regne Unit del 14%).⁽¹⁵⁾ Un estudi australià realitzat sobre una mostra de 3.000 adults, emprant mètodes epidemiològics rigorosos i actualitzats, va estimar una prevalença instantània del 25,5%, una prevalença periòdica als 6 mesos del 64,6% i una prevalença al llarg de la vida del 79,2%.⁽¹⁶⁾ Altres estudis europeus han comunicat una prevalença anual entre el 15-45% i una prevalença al llarg de tota la vida que arriba en algun estudi fins al 85%.⁽¹⁷⁻²⁰⁾ L'estudi australià abans mencionat va estimar retrospectivament la incidència anual d'un primer episodi de lumbàlgia en un 8,0%. Aquestes dades suggereixen que la lumbàlgia és una entitat molt freqüent a la població australiana, on quatre de cada cinc adults experimentaran un episodi al llarg de la seva vida i aproximadament un de cada 12 experimentarà un nou episodi (recurrència) en el decurs del següent any.

El principal estudi portat a terme al nostre medi (estudi EPISER), un estudi sobre la prevalença de malalties reumàtiques a la població adulta espanyola

realitzat sobre una mostra aleatòria de 2.192 adults major de 20 anys, va estimar una prevalença instantània de lumbàlgia del 14,8% (IC95% 12,2-17,4). La probabilitat de patir al menys un episodi de lumbàlgia va ser del 44,8% (IC95% 39,9-49,8). La prevalença estimada de lumbàlgia crònica va ser del 7,7% (IC95% 4,7-11,6).⁽²¹⁾

La prevalença de lumbàlgia en nens és baixa (1-6%) però augmenta ràpidament en adolescents (18–50%), presentant un pic entre els 55 i 65 anys.

La lumbàlgia és més freqüent en països amb una elevada renda per càpita i amb polítiques més liberals i major grau de cobertura social i de compensacions per baixa laboral i discapacitat, com per exemple, Alemanya, Suècia o Bèlgica. Així, a Suècia, el nivell de cobertura per lumbàlgia incapacitant és del 100%, en comparació amb els EEUU, on la cobertura oscil·la entre el 0 i el 80%. Al 1987, el percentatge de treballadors amb una baixa laboral per dolor d'esquena a Suècia va ser del 8% font el 2% als EEUU. La baixa laboral va durar, en promig, 40 dies a Suècia front a només 9 dies als EEUU.⁽²²⁾ Aquests resultats han estat reproduïts per estudis fets a d'altres països.

Afortunadament, la majoria dels episodis de lumbàlgia mecànica es resolen favorablement en el curs de 2 a 4 setmanes, independentment del tractament rebut.^(23,24) No obstant, el percentatge de casos de lumbàlgia que persisteix més enllà dels 3 mesos (lumbàlgia crònica) se situa entorn al 15% en adults i 27% en els pacients d'edat avançada.⁽¹⁴⁾ Lamentablement per als individus que no es recuperen dins els tres primers mesos, el procés de recuperació és lent, amb un impacte significatiu sobre la funcionalitat i una important despesa sanitària (visites repetides i nombroses proves i tractaments).

Aquest concepte de classificació de la lumbàlgia segons la seva durada (aguda: <6 setmanes, subaguda: de 6 a 12 setmanes, i crònica: >12 setmanes) ha estat freqüentment qüestionat ja que la lumbàlgia és una patologia que tendeix a recórrer i la majoria de pacients experimentaran episodis múltiples.⁽²⁵⁾ Estudis epidemiològics han mostrat que, després d'un primer episodi de lumbàlgia, la proporció de pacients que experimenten una recurrència a l'any pot arribar a ser del 60%.⁽²⁶⁻²⁷⁾

El mal d'esquena crònic és la principal causa de dolor crònic persistent, sent més freqüent en pacients d'edat avançada, associant-se amb limitacions funcionals i dificultats per dur a terme les activitats de la vida diària. En els pacients amb lumbàlgia crònica (dolor que dura més de 3 mesos), entre el 25% i 60% dels casos presenten dolor persistent a l'any o més després de l'episodi inicial.

En els individus més joves de 50 anys, la lumbàlgia mecànica representa la causa més freqüent de discapacitat laboral i s'associa habitualment a una lesió en el treball, mentre que en els major de 50 anys ocupa el tercer lloc. El nombre promig de dies de baixa laboral per lumbàlgia al nostre medi és de 21,9 dies en un estudi que valorava el període entre 1993 i 1997.⁽²⁸⁾

Després de la grip, el dolor d'esquena constitueix el segon motiu de consulta al metge general als EEUU.⁽²⁹⁾ Al 2002, representava aproximadament el 2,3% de totes les visites mèdiques als EEUU.⁽³⁰⁾ Al Regne Unit, s'estima que cada any 1 de cada 15 persones adultes (equival a 2,6 milions) consulten al seu metge de capçalera per lumbàlgia.⁽³¹⁾ En un estudi de l'Organització Mundial de la Salut (OMS) sobre dolor crònic i benestar a l'atenció primària, el 22% dels pacients varen comunicar dolor persistent, dels quals un 48% reportaven mal d'esquena crònic, sovint associat amb malaltia psicològica i discapacitat.⁽³²⁾

El dolor lumbar és, per tant, un trastorn comú i invalidant a la societat occidental, que representa una gran càrrega econòmica en forma de costos directes com a resultat de la pèrdua de productivitat laboral i les despeses mèdiques, i també de costos indirectes. S'estima que el 28% de la població de treballadors industrials dels EEUU experimentaran un episodi de lumbàlgia incapacitant en algun moment de la seva vida (sobretot en treballs amb alta demanda física com la construcció i la indústria), i el 8% de tota la població treballadora. Això suposarà el 40% de totes les jornades laborals perdudes.⁽¹⁴⁾ A més, el nombre de persones amb discapacitat que reclamen compensacions econòmiques per causa de la lumbàlgia és també cada vegada major. Aquestes representaven entre el 9% i 26% de les reclamacions a les companyies asseguradores. Al Regne Unit, la lumbàlgia va ser la causa més freqüent d'absentisme laboral des de 1988 a 1989 i va costar més de

£11.000.000.000.⁽³³⁾ Als EEUU, un estudi va mostrar que els costos anuals de salut i incapacitat per càpita de la lumbàlgia varen ser comparables amb els de la cardiopatia i la diabetis a una empresa important dels EEUU.⁽³⁴⁾

Contràriament al que se sol pensar, la prevalença de lumbàlgia crònica augmenta.⁽³⁵⁾ Un estudi poblacional als EEUU que avaluava la tendència en la prevalença de lumbàlgia crònica al llarg d'un període de 14 anys (1992-2006) a l'estat de Carolina del Nord, emprant mètodes similars i definicions de lumbàlgia crònica idèntiques al llarg del temps, va detectar un increment alarmant en tots els grups de població analitzats (per sexe, edat i grup racial). L'increment global al llarg d'aquest període va ser del 162%. També varen observar un increment, encara que de menor magnitud, en la prevalença d'episodis aguts, la qual cosa fa pensar en una taxa cada vegada major de casos aguts que evolucionen cap a la cronicitat. Una possible explicació d'aquesta tendència observada és el canvi observat en aquest mateix període en la composició de la força laboral de l'estat de Carolina del Nord, amb un augment del percentatge de treballadors al sector de la construcció i la indústria, així com també un notable increment (més del doble) en la prevalença de depressió major entre la població.⁽³⁶⁾

Altres possibles explicacions de l'augment de freqüència de la lumbàlgia és el major coneixement i sensibilització social pels seus símptomes, en gran part deguda als mitjans de comunicació, i consegüentment, la major taxa de comunicació espontània o de respostes positives a les enquestes així com de demanda d'atenció mèdica.

L'augment important observat en la despesa sanitària (visites al metge generalista o als especialistes, proves complementàries, tractaments curatius i rehabilitadors, i cirurgia) per causa de la lumbàlgia feia pensar que, a banda d'un increment en la seva prevalença, cada vegada hi ha un major grau d'utilització de recursos sanitaris per part dels pacients amb lumbàlgia. No obstant, alguns estudis han mostrat que aquest increment en la utilització de recursos sanitaris, inclosa la cirurgia, és sobretot secundari a la prevalença creixent de lumbàlgia crònica i no pas a un major consum per part dels respectius pacients, al menys a Carolina del Nord.⁽³⁵⁾ En qualsevol cas, es

tracta d'una patologia amb uns costos totals estimats als EEUU entre 100 i 200 mil milions de dòlars anuals.⁽³⁷⁾ Als països europeus, s'estima que el maneig de la lumbàlgia genera cada any un cost equivalent a entre l'1,7% i el 2,1% del seu Producte Interior Brut.⁽³⁸⁻⁴⁰⁾

El tractament de la lumbàlgia inespecífica

El tractament de la lumbàlgia implica la participació d'una gran varietat de professionals de la salut, especialment a l'àmbit de l'atenció primària. Tot i que hi ha àmplies variacions entre països, en el procés hi intervenen metges generalistes, fisioterapeutes, massatgistes, quiropràctics, terapeutes ocupacionals i de l'exercici, cirurgians ortopèdics, neurocirurgians, reumatòlegs, metges especialistes en medicina física i rehabilitadora i d'altres. Això implica un grau de formació i ensinistrament molt variable i una predilecció per un tipus o altre de tècniques i intervencions, la qual cosa explicaria, en part, la gran variabilitat observada en el tractament d'aquesta patologia.

D'altra banda, són molt nombroses les intervencions disponibles per al tractament de la lumbàlgia. A banda dels medicaments per tractar els símptomes (principalment, analgèsics i antiinflamatoris, però també relaxants musculars, esteroides, antidepressius o antiepilèptics), hi ha també una àmplia varietat de tractaments no farmacològics conservadors i invasius que són emprats amb molta freqüència, mostrant una gran variabilitat entre els diferents professionals de la salut implicats en el maneig d'aquests pacients. Està clar que existeix una important pressió de la indústria i dels col·lectius professionals implicats en cada cas per fomentar l'ús de cadascuna d'aquestes tecnologies. No obstant, moltes d'aquestes teràpies estan basades en una evidència molt feble, sense que hagin estat validades mitjançant mètodes científicament sòlids i hagin demostrat ser eficaces, segures, efectives i eficients.^(25,41) Així, depenent de la capacitat de pressió dels grups involucrats en l'ús de cada tecnologia, és molt variable el seu ús en els diferents àmbits. En alguns casos, l'ús d'aquests tractaments s'ha expandit molt ràpidament en indicacions que no han estat prou validades, existint encara molta incertesa sobre la seva eficàcia i seguretat. Donat que moltes d'aquestes intervencions es promouen a l'àmbit de la medicina privada, s'han donat casos d'un *marketing* abusiu i enganyós. És

per aquest motiu que l'augment de costos de l'atenció de la lumbàlgia no s'ha vist acompanyat per una millora proporcional en els resultats clínics o de la taxa discapacitat.

Davant d'aquesta situació, i també de la pràctica absència de guies de pràctica clínica per al maneig de la lumbàlgia crònica (les més de 30 guies preexistents tractaven sobre lumbàlgia aguda, quant el major impacte el provoca la forma crònica), la Comissió Europea va impulsar l'elaboració d'una exhaustiva guia de pràctica clínica basada en l'evidència científica sobre la prevenció i el tractament de la lumbàlgia inespecífica aguda i crònica (Guia COST B13) des de la perspectiva de l'atenció primària.⁽¹⁾ Aquesta guia, publicada l'any 2004, revisa una gran varietat d'intervencions farmacològiques, no farmacològiques conservadores i invasives, i basa les seves recomanacions en l'evidència disponible a favor o en contra de cadascuna d'elles. Aquesta guia pretén ajudar a prendre decisions ben informades basades en l'evidència científica, millorar la qualitat i els resultats de l'atenció mèdica, promoure un ús més racional i eficient dels recursos disponibles, així com identificar llacunes de coneixement que permetin prioritzar els futurs esforços d'investigació.

Concretament, la guia COST B13 revisa fins a 13 intervencions per la lumbàlgia aguda i 23 per la lumbàlgia crònica, totes elles utilitzades en major o menor mesura a nivell europeu, la qual cosa dóna una idea de la gran dispersió existent en aquest àmbit i la necessitat d'establir uns criteris basats en la revisió sistemàtica i síntesi de l'evidència científica que ajudin a identificar i promoure l'ús dels tractaments efectius i segurs alhora que evitar aquells altres que no han demostrat la seva utilitat. L'elevat nombre d'intervencions que no es recomanen a la guia posa de manifest la necessitat d'aquesta tasca de revisió i síntesi així com de la seva actualització periòdica a fi d'incorporar els nous estudis que puguin reforçar o modificar les recomanacions prèvies.

Taula 2. Recomanacions terapèutiques de la Guia COST B13.

Lumbàlgia aguda no específica	Lumbàlgia crònica no específica
Tractaments o intervencions no farmacològiques avaluades per la guia: 1)	Tractaments o intervencions no farmacològiques avaluades per la guia: 1)

<p>informació i educació, 2) repòs al llit, 3) consell de romandre actiu, 4) exercicis, 5) fàrmacs (analgèsics, antiinflamatoris i relaxants musculars), 6) infiltracions amb corticoides, 7) manipulació espinal, 8) escoles de l'esquena (<i>"back schools"</i>), 9) tractament conductual, 10) tracció, 11) massatges, 12) TENS (estimulació nerviosa elèctrica transcutània) i 13) programes de tractament multidisciplinaris a l'àmbit laboral.</p> <p>Altres tractaments no avaluats: acupuntura, herbes medicinals, teràpia interferencial i calor superficial.</p>	<p>teràpia interferencial, 2) laserteràpia, 3) suports lumbar, 4) ones de xoc, 5) ultrasons, 6) termoteràpia, 7) tracció, 8) TENS, 9) exercicis, 10) manipulació/mobilització, 11) massatge, 12) escoles d'esquena, 13) intervencions educacionals breus, 14) teràpia cognitiva-conductual, 15) tractament multidisciplinar (bio-psico-social), 16) acupuntura, 17) injeccions i bloquejos nerviosos, 18) neurorreflexoteràpia, 19) PENS (estimulació nerviosa elèctrica percutània), 20) Procediments de denervació per radiofreqüència o electrotèrmics (IDET), 21) estimulació de la medulla espinal, 22) cirurgia i 23) fàrmacs (antidepressius, relaxants musculars, antiinflamatoris, opioïds, antiepilèptics i pegats de capsaïcina).</p>
<p>Tractaments recomanats: la informació i educació; el consell de romandre actiu i continuar les activitats diàries, inclòs el treball si és possible; prescripció de fàrmacs per al dolor (preferiblement paracetamol) en determinades circumstàncies; la manipulació espinal per pacients que no aconsegueixen retornar a les seves activitats normals; i programes de tractament multidisciplinaris a l'àmbit laboral per treballadors de baixa més de 4-8 setmanes.</p>	<p>Tractaments recomanats: els exercicis supervisats com a tractament de primera línia, un tractament breu de manipulació/mobilització, les escoles d'esquena per aconseguir una milloria provisional (6 setmanes) del dolor i la capacitat funcional però no a llarg termini, la neurorreflexoteràpia o el PENS per aconseguir una milloria del dolor (si estan disponibles), i la cirurgia de fusió vertebral si després de 2 anys han fracassat els tractaments conservadors. També es recomana el tractament amb fàrmacs sota determinades circumstàncies.</p>
<p>Tractaments no recomanats: repòs al llit; exercicis; infiltracions amb corticoides; escoles de l'esquena (<i>"back schools"</i>); tractament conductual; tracció; massatges; TENS.</p>	<p>Tractaments no recomanats: teràpia interferencial; laserteràpia; suports lumbar; ones de xoc; ultrasons; termoteràpia; tracció; TENS; massatge; intervencions educacionals breus; teràpia cognitiva-conductual; tractament multidisciplinar (bio-psico-social); acupuntura; injeccions i bloquejos nerviosos; procediments de denervació per radiofreqüència o electrotèrmics (IDET); fàrmacs antiepilèptics.</p>

Pel que fa a la lumbàlgia crònica, és de notar les recomanacions que fa la guia COST B13 sobre la futura recerca. Aquestes són més extenses que no pas les recomanacions que es fan sobre la pràctica clínica, la qual cosa il·lustra les mancances o buits encara existents en aquest àmbit i la necessitat de portar a terme més estudis científics, especialment assaigs clínics.

Quant als aspectes metodològics generals, la guia recomana que els estudis sobre eficàcia/efectivitat que es facin en el futur siguin de qualitat, idealment amb un disseny d'assaig clínic controlat i aleatoritzat, i que incloguin anàlisis de costos i de risc-benefici. També es recomana que es facin estudis per determinar les modalitats d'aplicació i el perfil de professionals que obtenen millors resultats en aplicar les intervencions així com la identificació de poblacions específiques de pacients amb més probabilitat d'obtenir un benefici (desenvolupament d'eines que permetin una millor classificació dels pacients). Finalment, es recomana desenvolupar mesures (instruments i criteris) que siguin clínicament rellevants i sensibles, especialment pel que fa a la capacitat física i la funcionalitat, a fi de comprendre millor quin és l'impacte real dels tractaments sobre la salut dels pacients.

Quant a les modalitats específiques de tractament avaluades per la guia, aquesta inclou un seguit de recomanacions força detallades que afecten la quasi totalitat de les mateixes (veure taula 3).

Taula 3. Recomanacions per a la recerca de la Guia COST B13.

Fisioteràpia

Es necessita més investigació per avaluar els components específics dels tractaments habitualment utilitzats pels fisioterapeutes, mitjançant la comparació del seu ús aïllat i combinat. La combinació de certs tractaments físics 'passius' per alleujar el dolor simptomàtic amb altres tractaments més 'actius' dirigits a reduir la discapacitat (per exemple, massatges, compreses calentes o TENS, juntament amb la teràpia d'exercici) s'ha d'investigar més a fons. L'aplicació de principis cognitius i del comportament a la fisioteràpia, en general, necessita ser avaluat.

Tractament amb exercicis

L'efectivitat de tipus específics de tractament basat en exercicis ha de ser avaluada més a fons. Això inclou l'avaluació dels exercicis d'estabilització de la columna vertebral, els exercicis de McKenzie i altres règims populars d'exercici que s'utilitzen sovint però que s'han investigat de manera inadequada. S'ha d'investigar més sobre la intensitat òptima, freqüència i durada de l'exercici, així com la qüestió sobre l'exercici individual front al grupal. El "principi actiu" dels programes d'exercici és encara prou desconegut, el que requereix de més investigació per tal de permetre el desenvolupament i la promoció d'una major varietat de programes d'exercici de baix cost i efectius. L'aplicació de principis cognitius i del comportament per a la prescripció d'exercicis encara necessita ser avaluada.

Les escoles d'esquena i l'educació breu

El tipus d'assessorament i la informació proporcionada, el mètode de lliurament, i la seva eficàcia relativa necessiten ser avaluats, en particular pel que fa a les característiques del pacient i les seves creences/comportaments de base. Les característiques dels pacients que responen particularment bé a un mínim contacte (intervenció educativa breu) han de ser més investigades.

La teràpia cognitiva conductual

El valor relatiu dels diferents mètodes de teràpia cognitiva-conductual ha de ser avaluat. Els mecanismes d'acció subjacents també han de ser examinats per tal d'identificar subgrups de pacients que es beneficiaran més de la teràpia cognitiva-conductual i en els quals cal abordar determinats components de la persistència del dolor. S'han proposat predictors dels resultats del tractament conductual que necessiten ser avaluats, com ara la credibilitat del tractament, les etapes de canvi, les expectatives sobre els resultats, les creences (recursos d'afrontament, la por-evasió) i el catastrofisme. L'ús dels principis cognitius i del comportament per part de professionals no formats en psicologia clínica s'ha d'investigar per saber com aquests poden ser millor educats per aconseguir uns resultats millors.

Tractament multidisciplinari

El contingut òptim dels programes de tractament multidisciplinari requereix més investigació. S'ha de posar més èmfasi en la identificació del tractament adequat per al pacient adequat, especialment en relació amb l'extensió del tractament

multidisciplinari administrat. Això ha d'anar acompanyat d'anàlisi de cost-benefici.

Tractaments farmacològics

Hi ha molt poques dades en relació amb l'ús d'opioides (especialment opioides forts) per al tractament de la lumbàlgia crònica. Es requereixen més assaigs clínics. No hi ha estudis que hagin examinat els efectes dels AINEs a llarg termini en el tractament del dolor lumbar crònic. També es necessiten amb urgència estudis que incloguin l'avaluació de la funció. Es recomana fer estudis sobre l'eficàcia de paracetamol i metamicol (també, en comparació amb els AINE). El paper dels relaxants musculars, especialment en relació amb el seu ús prolongat, no està del tot clar i requereix més estudis.

Tractaments invasius

Cal investigar més sobre la selecció dels pacients, els procediments i les tècniques d'aplicació. En particular, calen estudis de més alta qualitat per examinar l'eficàcia de l'acupuntura, els bloquejos nerviosos i els procediments de la denervació per radiofreqüència i electrotèrmics.

Cirurgia

En primer lloc, cal avaluar els mètodes quirúrgics emergents en el context d'assaigs clínics controlats aleatoris d'alta qualitat, en què els millors tractaments conservadors disponibles basats en l'evidència (*gold standard*) serveixin de control. Els pacients que han fracassat a la cirurgia de l'esquena han de ser analitzats sistemàticament per tal d'identificar possibles indicacions quirúrgiques errònies o en el seu diagnòstic.

És evident, doncs, que encara hi ha moltes àrees en què cal investigar, ja sigui mitjançant estudis de recerca primària (assaigs clínics) com de recerca secundària o de síntesi (revisions sistemàtiques que ajudin a actualitzar els continguts i les recomanacions de les guies). Aquesta necessitat està molt ben expressada de forma resumida en les dues darreres recomanacions:

1. És possible que molts dels tractaments no recomanats per aquesta guia (per falta de proves o per dades conflictives sobre la seva efectivitat), puguin arribar a demostrar que són eficaços quan s'investiguin mitjançant assaigs clínics controlats i aleatoris d'alta

qualitat. Molts d'aquests mètodes de tractament s'utilitzen àmpliament, per tant, s'anima a què es realitzin estudis dissenyats acuradament per determinar si l'ús d'aquests mètodes està justificat

2. Els tractaments recomanats en aquesta guia no són eficaços per a tots els pacients amb lumbàlgia crònica. Calen més investigacions dirigides cap a la caracterització de la subpoblació de pacients amb lumbàlgia crònica a qui no ajuda cap dels tractaments considerats en aquesta guia.

Tractament de l'estenosi espinal

Aproximadament el 15% dels pacients amb lumbàlgia mecànica presenten una patologia (principalment hèrnia de disc lumbar i estenosi espinal) que es podria beneficiar potencialment de la cirurgia, encara que l'evidència disponible que indiqui que la cirurgia és millor que el tractament mèdic en aquestes patologies és encara limitada.⁽⁴²⁻⁴⁴⁾

Habitualment s'utilitza el terme estenosi espinal lumbar per descriure els pacients amb símptomes relacionats amb la reducció anatòmica de la mida del canal medul·lar a nivell lumbar. Es tracta, doncs, d'una definició de base anatòmica que, no obstant, no es correlaciona suficientment amb la intensitat dels símptomes i el grau d'incapacitat funcional dels pacients que la pateixen. De fet, es pot donar una estenosi espinal anatòmica molt important en pacients asimptomàtics. L'estenosi espinal es caracteritza per canvis degeneratius en els discs intervertebrals, el lligament flavum i/o les cares articulars, que estrenyen el canal espinal i produeixen una pressió sobre els elements neurals i el seu aport sanguini.⁽⁴⁵⁾ Els pacients habitualment presenten dolor lumbar, a les natges i l'extremitat inferior que augmenta a la bipedestació o al caminar i s'alleugera amb el descans (claudicació neurogènica). El diagnòstic es fa mitjançant estudis d'imatge que mostren l'estrenyiment del canal espinal.⁽⁴⁶⁾ Habitualment, esdevé simptomàtica durant la cinquena i sisena dècada de la vida, i afecta principalment els nivells L3-L4 i L4-L5. És la causa més freqüent de cirurgia espinal lumbar per damunt dels 65 anys d'edat.⁽⁴⁷⁾ L'envelliment s'associa amb una pèrdua d'alçada del disc que provoca un estretament del

foramen vertebral en sentit vertical, i que si és molt pronunciat pot també fer protruïr el lligament flavum dins el canal espinal.⁽⁴⁸⁾

Les indicacions de la cirurgia varien molt i la demostració d'una estenosi espinal radiològica sovint és asimptomàtica. Per això, és molt important demostrar una correlació clínica entre els símptomes i les imatges radiològiques d'estenosi.⁽⁴⁹⁾ L'impacte del dolor d'esquena, a la zona engonal i l'extremitat inferior i l'entumiment produït per l'estenosi espinal poden variar molt entre individus i és un factor clau a l'hora de decidir quin tractament s'ha d'administrar i quan. Per algunes persones, les intervencions conservadores (antiinflamatoris no esteroidals, exercici i teràpia física) funcionen bé per als símptomes, mentre que per altres el dolor els impedeix continuar les seves activitats diàries normals. En aquests individus, la cirurgia pot ser una opció raonable.⁽⁵⁰⁾ La cirurgia consisteix en la remoció d'os o d'elements de teixit tou que comprimeixen el sac dural o les arrels nervioses, l'ús d'empelt ossi i/o dispositius de fixació per tractar la inestabilitat espinal resultant, o ambdós. No obstant, l'evidència que mostri que la cirurgia és millor que la teràpia mèdica (tractament conservador) en aquesta indicació és escassa.

Malgrat la gran difusió de la cirurgia per al tractament de l'estenosi espinal, encara romanen incertes sobre els criteris diagnòstics, les indicacions específiques de la cirurgia, el procediment quirúrgic òptim i les característiques dels pacients associades amb un resultat quirúrgic favorable. La història natural de l'estenosi espinal és força desconeguda.

Un estudi sobre variabilitat geogràfica de la pràctica clínica als EEUU va mostrar que les taxes de cirurgia per estenosi espinal variaven 12 vegades entre unes zones i altres (de 0,29 per 1000 persones fins a 3,34 per 1000 persones).⁽⁴²⁾ La taxa va ser més del 50% superior que el promig nacional (0,99 per 1000) a 38 àrees i més del 50% inferior a 13 àrees. Aquesta variabilitat era molt superior a l'observada en d'altres procediments estudiats (per exemple, la taxa de cirurgia per fractura de maluc variava 2 vegades i la de prostatectomia radical 8 vegades).

Aquesta variabilitat reflecteix una manca de consens entre els clínics sobre els criteris diagnòstics i les indicacions d'aquest procediment. La taxa de cirurgia per fractura de maluc varia molt poc bàsicament perquè no hi ha discrepàncies respecte al seu diagnòstic i existeix un ampli consens clínic sobre el seu tractament òptim. Pel contrari, aquest consens no existeix per al tractament del càncer de pròstata precoç. Tampoc existeix una sòlida evidència que afavoreixi el tractament mèdic o quirúrgic de l'estenosi espinal. En absència de dades clíniques prou concloents derivades d'estudis de qualitat que puguin orientar la presa de decisions, és de preveure una alta variabilitat de la pràctica clínica. Així, els clínics poden diferir en la seva definició del què significa fracàs de les mesures conservadores, en els seus criteris radiològics per definir la presència de la malaltia, o en la manera com interpreten els símptomes i signes de l'exploració física.

La gran variabilitat observada indica que existeix un ampli marge de discrecionalitat en la presa de decisions sobre el tractament de l'estenosi espinal, en gran part degut a la incertesa sobre l'eficàcia relativa dels tractaments mèdics/conservadors i quirúrgics disponibles. La manca d'estudis de qualitat amb un disseny adequat per evitar els biaixos i que utilitzin mesures de resultat sensibles, fiables i vàlides que assegurin que els tractaments són avaluats d'una manera rellevant per als pacients, dificulta la presa de decisions.

La North American Spine Society (NASS) va elaborar una guia de pràctica clínica per al tractament de l'estenosi espinal degenerativa publicada l'any 2007.⁽⁵¹⁾ Es va arribar al consens de que a llarg termini, entre els pacients amb dolor lleu o moderat i tractats inicialment amb intervencions mèdiques/conservadores, aproximadament el 20-40% finalment necessitaran cirurgia. Dels pacients que no requereixen intervenció quirúrgica, el 50-70% tindran una milloria del seu dolor, encara que l'evidència disponible es reconeixia escassa. També es va concloure que en els pacients amb símptomes greus d'estenosi espinal lumbar, la cirurgia descompressiva sola és eficaç en el 80% del casos, i que el tractament mèdic aïllat és eficaç aproximadament en el 33% dels casos.⁽⁵¹⁾

En conclusió, la cirurgia no és la resposta definitiva per a tothom que presenti dolor d'esquena persistent. Ofereix un tractament específic per problemes anatòmics específics associats amb símptomes específics. Si no hi ha un dèficit neurològic important, els pacients amb discos herniats, espondilolistesi o estenosi espinal no necessiten una cirurgia, tot i que els procediments quirúrgics apropiats poden proporcionar un alleujament de dolor. En aquelles situacions, les decisions haurien de ser fetes conjuntament pels pacients i els seus metges ben informats. No obstant, malgrat les assercions de que la cirurgia és només un últim recurs, o que és utilitzada més selectivament que en el passat, la taxa de cirurgies d'esquena ha augmentat molt ràpidament en les darreres dècades, fins i tot després d'ajustar per l'edat de la població.⁽⁵²⁾ Els conflictes d'interès econòmic en la Medicina han tingut un impacte significatiu en la investigació i en la pràctica clínica relacionades amb la cirurgia de l'esquena. Poden determinar la realització de cirurgies innecessàries, de vegades molt agressives, així com la promoció de tecnologies poc avaluades en el mercat.⁽⁵³⁻⁵⁵⁾

2.2.- L'avaluació dels tractaments i les intervencions terapèutiques

La medicina basada en l'evidència i els assaigs clínics

La medicina basada en l'evidència (MBE) s'ha definit com l'ús conscient, explícit i judiciós de la millor evidència científica disponible en la presa de decisions sobre l'atenció sanitària dels pacients.⁽⁵⁶⁾ La pràctica de la MBE significa integrar la competència clínica individual amb la millor evidència clínica externa disponible a partir de la investigació sistemàtica (fonamentalment, estudis ben dissenyats centrats en els pacients i les seves necessitats) per tal de saber, d'una banda, entendre els resultats dels estudis i tenir una expectativa raonable dels beneficis que poden aportar els tractaments avaluats i, d'altra banda, conèixer les especificitats del pacient individual i tenir un criteri per decidir quan i com cal aplicar aquests tractaments.

Suposa un acostament a la presa de decisions en la qual els clínics usen la millor evidència disponible, en consens amb el pacient, per decidir quina serà l'opció més satisfactòria.

Quan es tracta de seleccionar un tractament, l'assaig clínic resulta el millor disseny per avaluar objectivament i de forma fiable els seus efectes, tant els beneficis com els riscos que comporten. L'assaig clínic ofereix la millor qualitat d'informació o el nivell més alt d'evidència possible, per damunt d'altres alternatives com són els estudis quasi-aleatoris, els estudis controlats no aleatoris, els estudis de cohort prospectiva o les sèries de casos. Les característiques intrínseques del seu disseny (fonamentalment, la comparació entre diferents opcions de tractament assignades de forma aleatòria o per atzar, comparades de manera prospectiva i simultània en el temps, aplicades sota unes mateixes condicions i amb l'intent d'avaluar el seu impacte en els pacients de la forma més objectiva possible mitjançant estratègies d'emascament o cegament de les intervencions) redueixen les possibilitats de biaix, oferint estimacions més fiables dels efectes del tractament i, per tant, permetent realitzar inferències causals que redueixin la incertesa.

Per això, l'assaig clínic constitueix el patró d'or (*gold standard*) per a validar la utilitat de les intervencions sanitàries, sigui quina sigui la seva naturalesa (intervencions farmacològiques o d'altre tipus). Això no significa que altres tipus de disseny no puguin proporcionar-nos informació d'utilitat sobre els tractaments o intervencions (per exemple, per avaluar la reproductibilitat dels seus efectes en un entorn determinat o en un tipus concret de pacients, per avaluar els seus efectes a llarg termini o per identificar efectes adversos greus però poc freqüents). No obstant, quan la seva utilitat en termes d'efectes beneficiosos és encara incerta, les seves potencials conseqüències rellevants sobre la salut dels pacients (per exemple, risc de toxicitat o de complicacions greus) i/o el seu cost elevat, resulta imprescindible disposar d'una informació que sigui veraç abans de poder prendre una decisió sobre el seu ús.

La normativa internacional que regula l'autorització i comercialització dels medicaments d'ús humà exigeix que s'aportin proves sobre la seva eficàcia i seguretat com a requisit previ a la seva autorització i posterior comercialització

per part de les agències reguladores. Especialment a partir de la catàstrofe de la talidomida als anys seixanta, la regulació a nivell internacional ha establert uns criteris ben definits cada cop més exigents, definint etapes successives en el desenvolupament d'un nou fàrmac, iniciant amb estudis preclínic en cèl·lules, teixits o animals fins arribar als estudis clínics en humans de fase I fins a la IV. Als darrers anys, aquestes normatives han anat convergint a nivell internacional mitjançant directrius d'harmonització, conegudes com a International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). Tot i que la normativa actual no resol tots els problemes i necessitats que es presenten des de la perspectiva dels decisors en salut (per exemple, l'exigència de comparacions directes entre alternatives actives que competeixen per una mateixa indicació, la selecció de variables de resultat clínicament rellevants o la transparència en la difusió del resultats de tots els estudis autoritzats), no obstant ha creat una cultura d'avaluació cada vegada més exigent i estandarditzada pel que fa als fàrmacs. Més darrerament, aquestes mateixes normes s'han estès també als dispositius sanitaris (Real Decreto 1591/2009, de 16 de octubre, por el que se regulan los productos sanitarios). No obstant, com alguns autors ja han assenyalat, les exigències d'aquest sistema no han estat al mateix nivell que per als fàrmacs, permetent la comercialització de dispositius o tecnologies que no han estat suficientment avaluades.⁽⁵⁷⁾ No hi ha regulacions pel que fa a d'altres tipus d'intervencions que no impliquin l'ús de fàrmacs o dispositius (per exemple, exercicis, intervencions educacionals o psicològiques o cirurgia), molt freqüents en l'àmbit del tractament de la lumbàlgia. A més, aquestes normatives no intervenen en el procés posterior a la difusió i comercialització de les tecnologies sanitàries (vigilància post-implantació, adequació d'ús, etc.).

Les revisions sistemàtiques

A mesura que han proliferat els assaigs clínics i el volum d'informació s'ha fet difícil de gestionar (problemes per a la localització, anàlisi i integració o síntesi de tota la informació disponible), les revisions sistemàtiques han anat ocupant un lloc cada vegada més prominent en el procés d'avaluació i en la presa de decisions respecte als tractaments o intervencions en salut, tant al nivell clínic

individual com de les autoritats responsables de la prestació o de la gestió dels serveis de salut a nivell comunitari. A més, les revisions sistemàtiques són la base per a la realització de guies de pràctica clínica, que recentment han estat definides com a *“documents informatius que inclouen recomanacions dirigides a optimitzar l’atenció dels pacients en base a una revisió sistemàtica de l’evidència i a l’avaluació dels beneficis i riscos de les diferents opcions alternatives en l’atenció a la salut”*.⁽⁵⁸⁾

Avui més que mai existeixen motius importants que fan necessari gestionar correctament el coneixement. En primer lloc, patim un excés d’informació; s’ha passat en tan sols uns pocs anys de no disposar de fonts d’informació a nedar en una sobreabundància que ens satura i que pot ser contraproduent. Es calcula que s’han realitzat centenars de milers d’assaigs clínics aleatoritzats sobre diferents aspectes de l’atenció sanitària des de 1948, any en què es va publicar el primer d’aquests estudis, i s’espera que la tendència continuï sent creixent. En segon lloc, disposem a l’actualitat d’un número inesgotable de fonts d’informació disponibles, no sempre accessibles. Així, es publiquen anualment més de 25.000 revistes biomèdiques a tot el món, de les quals 5.000 i un total de 21 milions d’articles es troben indexats només a MEDLINE, una entre moltes bases bibliogràfiques disponibles. Un metge que intenti mantenir-se al dia només en la seva especialitat hauria de llegir, teòricament, més de 100 articles complets cada setmana. D’altra banda, el temps disponible per la cerca d’informació i l’estudi és cada vegada més limitat. Com a conseqüència, els coneixement del metge (o de qualsevol altre professional de la salut) es deterioren ràpidament després de finalitzar el seu període de formació i d’ensinistrament en la seva especialitat. En tercer lloc, la qualitat de les diverses fonts d’informació és molt desigual, de manera que el problema al qual ens enfrontem no és només d’excés sinó també de qualitat de la informació, atès que sovint és deficient. Per últim, els continguts de les fonts d’informació queden obsolets ràpidament, la qual cosa suposa una limitació addicional dels sistemes actuals. Per tant, necessitem disposar d’eines que ens permetin accedir a la informació adequada en termes de quantitat, qualitat i actualitat.

Les revisions ens ofereixen la possibilitat d'estar ben informats en el menor temps possible. Precisament, una de les maneres més eficients de trobar l'evidència científica a la que fa referència la medicina basada en l'evidència és buscar una revisió sistemàtica de tots els assaigs clínics aleatoritzats (o en la seva absència, dels millors estudis disponibles) sobre els efectes de l'atenció sanitària. Per tant, necessitem imperiosament revisions sistemàtiques per integrar de forma eficient tota la informació vàlida (beneficis, riscos i costos) i proporcionar una base objectiva per a prendre decisions de manera racional. Les revisions sistemàtiques estableixen quan els efectes de l'atenció sanitària són consistents i quan poden aplicar-se els resultats de la investigació als diferents grups de població o tipus de pacients. L'ús de mètodes explícits i sistemàtics en les revisions limita el biaix (error sistemàtic) i redueix els efectes trobats per atzar, de manera que proporciona resultats més fiables sobre els quals treure conclusions i prendre decisions. La metanàlisi, és a dir, l'ús dels mètodes estadístics per a combinar i resumir els resultats d'estudis independents entre sí, poden proporcionar estimacions més precises sobre els efectes de l'atenció sanitària que les derivades dels estudis individuals inclosos en una revisió.

Les revisions sistemàtiques són investigacions científiques en sí mateixes (recerca secundària o recerca sobre la recerca), amb mètodes específics preestablerts a un protocol per dur a terme l'assemblatge dels estudis originals, que sintetitzen els resultats d'aquests. Les característiques bàsiques que millor defineixen una revisió sistemàtica es podrien resumir en les següents: a) són síntesis i anàlisis de la informació amb un enfocament pràctic; b) es basen en la millor evidència científica disponible; c) formulen preguntes clarament definides, i d) utilitzen mètodes sistemàtics i explícits per identificar i seleccionar els estudis pertinents, avaluar-los críticament, extraure les dades d'interès i analitzar-los. Conseqüentment, les revisions sistemàtiques pretenen ser: a) rigoroses, quant als estudis inclosos (amb criteris de qualitat, etc.); b) informatives, és a dir, enfocades cap a problemes reals, tractant de contestar una pregunta clarament delimitada o específica, i idealment analitzant i presentant les dades de la forma que millor ajudi a la presa de decisions; c) exhaustives: el seu objectiu és identificar i utilitzar la major quantitat possible

d'informació pertinent, sense introduir biaixos (de publicació, de selecció, etc.), i d) explícites, ja que tots els mètodes utilitzats a la revisió han d'escriure's amb el detall suficient.

2.3.- Justificació de la tesi per compendi d'articles

Aquesta tesi, que es presenta per la via de compendi de publicacions, es basa fonamentalment en quatre revisions sistemàtiques publicades sobre l'eficàcia i seguretat de quatre intervencions no farmacològiques que han estat desenvolupades per al tractament de diverses patologies de raquis (lumbàlgia inespecífica en dues d'elles, lumbàlgia d'origen discogènic en una altra i estenosi espinal en l'altra).

L'interès inicial de dur a terme aquestes revisions sorgeix arran de l'obtenció de dues beques del Instituto de Salud Carlos III (ETS expedient 03/1003 y FIS expedient PI030908) per avaluar diferents intervencions no farmacològiques i quirúrgiques aplicades en l'àmbit de les patologies de l'esquena. Les intervencions seleccionades ho varen ser per la controvèrsia existent sobre la seva potencial utilitat, alhora que es constatava un creixent ús de les mateixes al nostre àmbit. Les revisions que aquí es presenten són les que han estat publicades a dia d'avui en revistes indexades. Addicionalment, a l'annex s'inclou una cinquena revisió que ha estat recentment completada i enviada a publicació a una revista indexada (pendent de la decisió editorial).

A banda de l'objectiu específic de cadascuna de les revisions (avaluació de l'eficàcia i seguretat de diverses intervencions), la seva compilació en aquesta tesi pretén il·lustrar amb exemples concrets la potencial utilitat de les revisions sistemàtiques així com també les limitacions habituals en el disseny, execució i difusió dels estudis primaris (assaigs clínics) que avaluen intervencions no farmacològiques. Com alguns autors ja han assenyalat, sovint la realització d'aquest tipus d'estudi per avaluar intervencions no farmacològiques suposa uns reptes majors que els que afronta la investigació amb fàrmacs i que, sovint, determinen una inferior qualitat dels estudis realitzats, si no la 'impossibilitat' pràctica per dur-los a terme.⁽⁵⁹⁻⁶⁰⁾

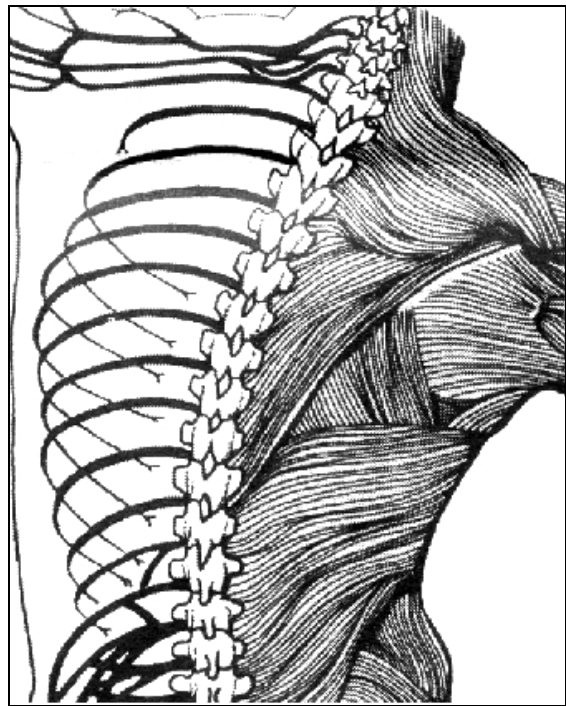
Totes les revisions han estat realitzades per un equip multidisciplinari que inclou també experts clínics en l'àrea de l'esquena. L'autor de la tesi hi ha participat en totes elles com a *metodòleg*. Per aquest motiu, l'èmfasi d'aquest document no pretén recaure en els aspectes més clínics dels temes que s'avaluen sinó més específicament en els metodològics relacionats amb la recerca.

En consonància amb això, i per tal de complementar la informació d'aquesta tesi, als annexos s'inclouen tres publicacions recents de l'autor de la tesi de caire metodològic, dues sobre la metodologia de les revisions sistemàtiques (un article i un capítol de llibre) i l'altra sobre les directrius actuals recomanades per les principals revistes biomèdiques sobre la publicació de revisions sistemàtiques i metanàlisis (PRISMA), a fi i efecte d'aclarir alguns aspectes de les revisions incloses a la tesi que no hagin estat suficientment explicats.

La cinquena i darrera publicació d'aquest compendi consisteix en un estudi descriptiu sobre la recerca portada a terme a Espanya sobre les patologies mecàniques del raquis. El seu interès va sorgir de la percepció de que no hi havia una correspondència entre l'impacte social i sanitari del problema que suposa les patologia de l'esquena (tal com s'ha esmentat a la introducció) i els recursos públics destinats a la recerca dels diferents aspectes d'aquesta malaltia, inclosa però no limitada a l'avaluació de les intervencions preventives o terapèutiques.

Donat que la recerca sobre les intervencions terapèutiques abarca, però no es limita, a la demostració de la seva eficàcia sinó que continua amb d'altres aspectes sovint no prou investigats, alguns dels quals tenen a veure amb la seva posterior implantació i difusió en el sistema de salut, s'ha cregut interessant afegir als annexos un altre article centrat en aspectes controvertits de la recerca en aquest àmbit de les patologies de l'esquena. Aquest article desenvolupa algunes idees que s'han emprat de forma més sintètica a la discussió d'aquesta tesi.

3. Objectius



3.- OBJECTIUS

3.1.- Objectius generals

Els objectius generals de la tesi són:

- 1.- Aprofundir sobre la utilitat, limitacions i reptes per dur a terme l'avaluació de l'eficàcia i seguretat de les intervencions no farmacològiques per al tractament de les patologies de l'esquena mitjançant la metodologia de les revisions sistemàtiques.
- 2.- Analitzar les limitacions i principals reptes metodològics per dur a terme assaigs clínics que avaluen intervencions no farmacològiques en aquest àmbit.
- 3.- Reflexionar sobre les deficiències en el procés d'aprovació i introducció d'aquestes intervencions al sistema de salut i fer algunes propostes de millora.

3.2.- Objectius específics

Els objectius específics de la tesi són els propis de cadascuna de les publicacions que la conformen.

Quant a les revisions sistemàtiques:

- 1.- Avaluar l'eficàcia i seguretat de la neuroreflexoteràpia (NRT) per al tractament de la lumbàlgia no específica en pacients adults. De forma secundària, es volia avaluar també l'efectivitat, comparant l'eficàcia de la NRT respecte d'altres intervencions convencionals en la mateixa patologia.
- 2.- Avaluar l'eficàcia, efectivitat i seguretat de les tècniques d'electrotermocoagulació intradiscal percutànies (IDET i PIRFT) per la lumbàlgia d'origen discogènic.

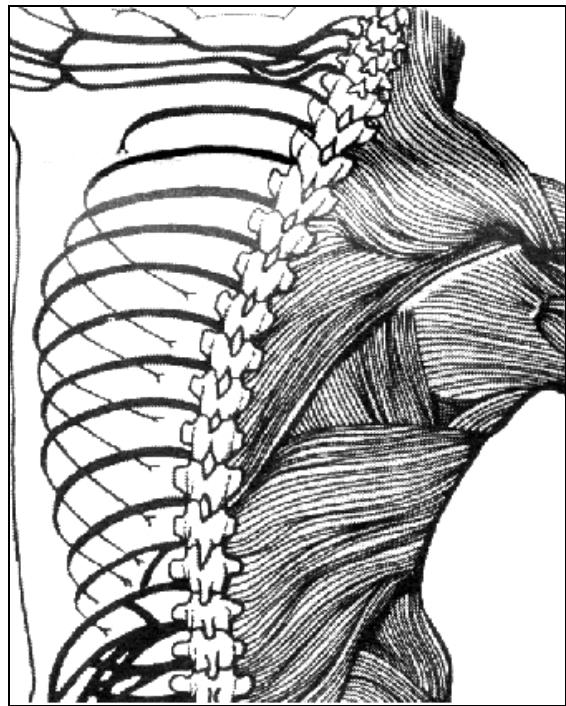
3.- Avaluar l'efectivitat de la cirurgia en comparació amb el tractament conservador sobre el dolor, discapacitat i qualitat de vida en l'estenosi lumbar simptomàtica.

4.- Avaluar l'eficàcia, seguretat, efectivitat i cost-efectivitat de les teràpies basades en ultrasons i ones de xoc per la lumbàlgia.

Quant a l'estudi de la recerca sobre patologies mecàniques del raquis duta a terme a Espanya:

5.- Descriure la recerca espanyola sobre patologies mecàniques del raquis, el seu import, fonts de finançament i la seva productivitat.

4. Mètodes



4.- MÈTODES

Els mètodes de la tesi són els propis de cadascuna de les publicacions que la conformen.

► Quant a les primeres quatre publicacions, es tracta de **revisions sistemàtiques**.

Una revisió sistemàtica és la revisió d'una pregunta clínica (en aquest cas, una pregunta terapèutica) formulada clarament, que utilitza mètodes sistemàtics i explícits per a identificar, seleccionar i avaluar críticament la investigació rellevant (els estudis originals pertinents), així com per a obtenir i analitzar les dades dels estudis inclosos a la revisió. Poden utilitzar o no mètodes estadístics (metanàlisi) per analitzar i resumir els resultats dels estudis inclosos.

Totes elles inclouen les següents seccions:

1.- Antecedents: consisteix en una breu justificació de la importància i necessitat de dur a terme la revisió. En resum, una revisió es justifica per la importància del problema de salut de que tracta i la controvèrsia quant a la/es intervenció/ns que s'analitzen.

2.- Objectiu/s de la revisió: en totes elles es tracta de l'avaluació de l'eficàcia, l'efectivitat i seguretat de diverses intervencions terapèutiques.

3.- Criteris de selecció dels estudis: condicions que han d'acomplir els estudis de recerca original per ser seleccionats i inclosos a la revisió. Concretament, els estudis han de complir uns requisits que tenen a veure amb: i) la població d'estudi, ii) les intervencions avaluades, iii) les comparacions que són d'interès per la revisió (placebo, intervenció simulada o 'no fer res' per avaluar eficàcia, o una altra intervenció activa per avaluar efectivitat), iv) els resultats que són d'interès (segons una jerarquitització establerta *a priori* en funció de la seva importància clínica per als clínics i els pacients), i v) el tipus d'estudi (disseny).

Els estudis se seleccionen si compleixen tots els criteris abans esmentats, i es descarten si no en satisfan almenys un d'ells. En conseqüència, la decisió

sobre la selecció dels estudis de recerca primaris és completament independentment dels seus resultats.

4.- Estratègia de cerca: la revisió, si ha de ser sistemàtica i exhaustiva, ha d'explicitar les bases bibliogràfiques que s'han emprat a la cerca (obligatòriament, Medline, Embase i el Registre Central Cochrane d'Estudis Controlats –CENTRAL-, en ocasions, altres bases més específiques segons l'àrea temàtica de la revisió; l'estratègia dissenyada (paraules clau i termes MeSH); restriccions idiomàtiques (en una revisió sistemàtica no n'hi ha d'haver); altres estratègies emprades, com seguiment de les referències o contacte amb experts.

5.- Mètodes de la revisió: cal explicitar com (i qui) s'han aplicat els criteris de selecció d'estudis, com (i qui) s'ha portat a terme l'avaluació de la qualitat dels estudis (a les revisions d'aquesta tesi s'han aplicat els criteris de Van Tulder adoptats pel grup de revisió Cochrane sobre esquena)⁽⁶¹⁾, com (i qui) s'ha dut a terme l'extracció de dades dels estudis seleccionats i l'anàlisi de resultats (especificant els models estadístics emprats per la combinació de resultats, si procedeix, l'anàlisi de l'heterogeneïtat, i les mesures de l'efecte emprades per variables dicotòmiques i contínues). Donat que moltes de les decisions que cal prendre en el procés d'una revisió sistemàtica poden tenir algun grau de subjectivitat o error (per exemple, selecció d'estudis, extracció de dades o avaluació de la qualitat dels estudis), cal que tots aquests processos siguin realitzats per dos revisors de forma independent i, en el cas d'haver discrepància, aquesta ha de ser resolta per consens.

6.- Descripció dels estudis: la revisió ha de proporcionar una descripció dels estudis seleccionats, resumida en forma de taula, per tal que els lectors puguin copsar a cop d'ull les característiques bàsiques de les poblacions d'estudi, els mètodes emprats, la modalitat específica d'intervenció i control aplicades i les variables de resultat (amb la seva definició o mètodes de mesura emprats) analitzades a cada estudi.

7.- Qualitat dels estudis: una part molt important de la revisió consisteix en avaluar el risc de biaix dels estudis seleccionats, a partir de la valoració de les

dimensions metodològiques essencials del disseny i execució d'un assaig clínic. Bàsicament, aquestes tenen relació amb l'assignació dels tractaments, l'ocultació de la seqüència de les assignacions, l'emascament de les intervencions als participants i/o als avaluadors dels resultats, la similitud de les característiques basals dels grups d'estudi, les pèrdues i abandonaments i el seu maneig a l'anàlisi dels resultats, l'aplicació d'un mateix cronograma de les avaluacions, o l'adequat compliment terapèutic.

8.- Resultats: d'una banda, hi ha els resultats obtinguts d'aplicar la cerca bibliogràfica i, d'altra banda, els resultats de l'anàlisi de l'impacte (efecte) dels tractaments en els subjectes d'estudi. Quant als primers, la revisió ha d'informar del flux d'estudis al llarg del procés de cerca, reportant el número d'estudis inicialment detectats, els motius d'exclusió i el número d'estudis finalment seleccionats. Quant als segons, la revisió ha d'informar de manera clara i concisa quina és la direcció i magnitud de l'efecte dels tractaments sobre cada variable de resultat analitzada per la revisió, presentant aquesta informació de forma resumida quantitativament mitjançant la combinació de resultats (metanàlisi) o bé de forma qualitativa quan no es compleixen les condicions per dur a terme aquesta combinació de dades.

9.- Discussió dels resultats: a partir dels resultats, i evitant en tot el possible transferir els valors i preferències personals, cal interpretar els resultats cenyint-se estrictament als mètodes i criteris prefixats al protocol de la revisió. També cal posar en context els resultats, especialment des del punt de vista de la seva rellevància clínica.

10.- Conclusions: finalment, els autors han d'arribar a emetre les seves conclusions, les quals tindran implicacions a dos nivells diferents i complementaris: 1) les implicacions per a la pràctica clínica serien aquelles recomanacions que es poden derivar d'uns resultats prou clars (a favor o en contra) de l'ús de la intervenció analitzada, amb graus variables de certesa; 2) les implicacions per a la investigació, en canvi, identifiquen llacunes del coneixement o deficiències en la recerca prèvia, que justificarien la necessitat de dur a terme nous estudis de recerca focalitzats a suplir aquestes mancances i aportar nova informació rellevant.

11.- Referències: la revisió, com tota investigació formal, ha d'aportar les referències bibliogràfiques bàsiques, que han d'incloure sobretot la referències dels estudis inclosos sobre la qual aquesta es basa.

12.- Taules i figures: la revisió ha d'incloure taules i figures on, de forma resumida, es presenta la descripció dels estudis, el flux d'estudis al llarg del procés de revisió, l'avaluació de la seva qualitat o risc de biaix i els resultats en el cas d'haver-hi metanàlisi.

La metodologia general de les revisions sistemàtiques es pot consultar al Manual Cochrane de Revisions (en anglès, Cochrane Handbook for Systematic Reviews of Interventions), lliurement disponible al web: www.cochrane-handbook.org/. També es poden consultar les dues publicacions dels annexos 2 i 3 de caràcter divulgatiu sobre la matèria.

La metodologia més específica (especialment pel que fa als criteris de selecció i l'estratègia de cerca) cal consultar-la directament a les publicacions.

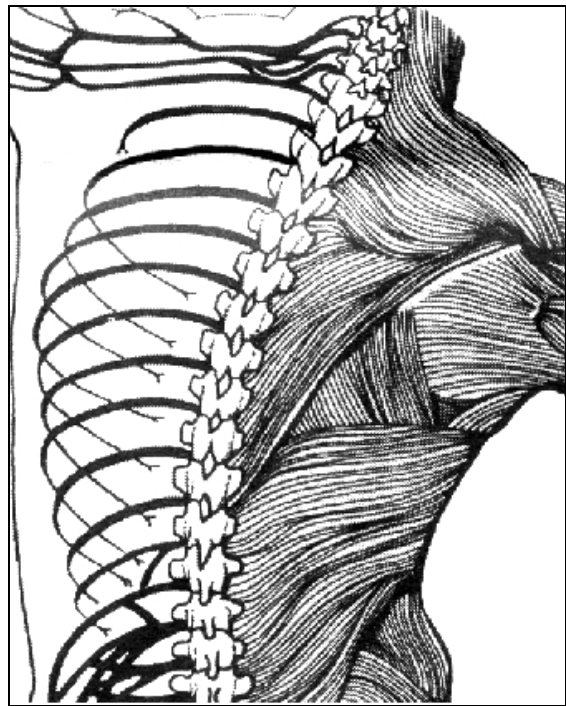
► Quant a l'estudi sobre la investigació espanyola de patologies mecàniques del raquis, es tracta d'un **estudi descriptiu de cohort retrospectiva** dels estudis portats a terme en el període 1992 a 2006 i de les seves publicacions derivades.

L'estratègia d'identificació dels estudis i les seves publicacions es va basar en enquestes complementades per cerques electròniques a bases bibliogràfiques.

Les dades analitzades consisteixen en la descripció dels estudis segons l'entitat finançadora, import atorgat, publicacions derivades, factor d'impacte i tipologia de recerca (bàsica o clínica –diferenciant entre assaigs clínics, revisions i altres).

Els detalls sobre la metodologia emprada poden consultar-se a l'article.

5. Resultats



5.- RESULTATS

5.1.- Publicacions presentades per aquesta tesi

Publicació nº 1: **Urrútia G, Burton AK, Morral Fernández A, Bonfill Cosp X, Zanolli G. Neuroreflexotherapy for non-specific low-back pain. Cochrane Database of Systematic Reviews 2004, Issue 2. Art. No.: CD003009. DOI: 10.1002/14651858.CD003009.pub2.**

Publicació nº 2: **Urrútia G, Kovacs FM, Nishishinya MB, and Olabe J. Percutaneous Thermocoagulation Intradiscal Techniques for Discogenic Low Back Pain. Spine 2007;32:1146–1154.**

Publicació nº 3: **Kovacs FM, Urrútia G, Alarcón JD. Surgery versus conservative treatment for symptomatic lumbar spinal stenosis: a systematic review of randomized controlled trials. Spine (Phila Pa 1976). 2011;36:E1335-E1351.**

Publicació nº 4: **Seco J, Kovacs FM, Urrútia G. The efficacy, safety, effectiveness, and cost/effectiveness of ultrasound and shock wave therapies for low back pain: A systematic review. Aceptado para publicación en Spine J (2011).**

Publicació nº 5: **Urrútia G, Bonfill X. Artículo especial. Análisis de la investigación española sobre dolencias del cuello y la espalda (1992–2006). Med Clin (Barc).2010;135(5):215–221.**

5.2.- Resum dels resultats

Publicació nº 1: *Urrútia G, Burton AK, Morral Fernández A, Bonfill Cosp X, Zanolli G. Neuroreflexotherapy for non-specific low-back pain. Cochrane Database of Systematic Reviews 2004, Issue 2. Art. No.: CD003009. DOI: 10.1002/14651858.CD003009.pub2.*

Factor d'impacte (2010): 6,186

Objectiu nº 1: Avaluar l'eficàcia i seguretat de la neuroreflexoteràpia (NRT) per al tractament de la lumbàlgia no específica en pacients adults. De forma secundària, es volia avaluar també l'efectivitat, comparant l'eficàcia de la NRT respecte altres intervencions convencionals en la mateixa patologia.

La cerca actualitzada fins a juliol de 2009 va permetre identificar tres assaigs clínics (AC) aleatoritzats, amb un total de 273 pacients amb lumbàlgia no específica, tots tres realitzats pel mateix investigador principal però amb equips de recerca diferents de l'estat espanyol. Dos AC es varen fer a l'àmbit de l'atenció primària i l'altre en un hospital universitari (servei de Reumatologia).

Dos AC varen comparar NRT amb placebo (NRT simulada) i l'altre va comparar NRT amb l'atenció convencional (a discreció del metge de família responsable del pacient). Aquest darrer estudi va consistir en un AC aleatoritzat per conglomerats (21 metges). El seguiment dels pacients va oscil·lar entre 30 y 60 dies. Dos AC varen fer un seguiment addicional a llarg termini (1 any) només dels pacients del grup d'intervenció.

Dos dels tres AC varen tenir un disseny adequat (baix risc de biaix) mentre que el tercer presenta algunes limitacions (per exemple, pèrdues no explicades, anàlisis dels casos vàlids -dades disponibles-, disseny obert i diferències en les característiques basals).

La població d'estudi va consistir fonamentalment en casos de lumbàlgia crònica i també (una minoria) subaguda, sense que fos possible diferenciar els resultats en ambdues subpoblacions.

Quant a l'eficàcia de la NRT (comparació vs NRT simulada), els dos AC varen mostrar una reducció significativa del dolor d'esquena i referit amb NRT al final de l'estudi (30 i 45 dies). La reducció del dolor amb NRT va ser d'una magnitud molt important (50%), amb un efecte d'inici molt immediat i sostingut. L'efecte placebo va ser només del 10%. Resultats molt similars es varen observar en les diferents mesures de dolor realitzades (dolor a la palpació i la flexió). Altres mesures també favorables a NRT varen ser una mesura global de la discapacitat per les activitats diàries, la capacitat funcional, dies de baixa laboral, dies de repòs al llit, consum d'analgèsics i efectes adversos. No es varen observar diferències significatives en la qualitat de vida.

Quant a l'efectivitat de la NRT comparada amb l'atenció mèdica habitual, la NRT va reduir significativament el dolor d'esquena i referit al final de l'estudi (60 dies). També es varen obtenir millors resultats de forma significativa en la incapacitat, durada de la baixa laboral i consum de recursos sanitaris. No es varen observar diferències significatives en la qualitat de vida.

L'estudi va incloure una anàlisi de cost-efectivitat que va resultar favorable a NRT.

En conclusió, la NRT sembla ser una intervenció eficaç, efectiva i segura per al tractament de la lumbàlgia crònica no específica. L'eficàcia és menys clara per als pacients amb lumbàlgia subaguda. No obstant, aquests resultats estan limitats a estudis realitzats per un petit nombre de terapeutes altament ensinistrats en aquesta tècnica. No es coneix la corba d'aprenentatge necessària per reproduir aquests resultats. Per tant, la NRT només es pot recomanar allà on aquesta tècnica està disponible i aplicada per part de personal altament ensinistrat.

Publicació nº 2: *Urrútia G, Kovacs FM, Nishishinya MB, and Olabe J. Percutaneous Thermocoagulation Intradiscal Techniques for Discogenic Low Back Pain. Spine (Phila Pa 1976). 2007;32:1146–54.*

Factor d'impacte (2007): 2,499

Objectiu nº 2: Avaluar l'eficàcia, efectivitat i seguretat de les tècniques d'electrotermocoagulació intradiscal percutànies (IDET *-intradiscal electrothermal annuloplasty-* i PIRFT *-percutaneous radiofrequency annuloplasty-*) per la lumbàlgia d'origen discogènic.

La cerca actualitzada fins a desembre de 2005 va permetre identificar 4 AC aleatoritzats i dos estudis controlats no aleatoris, amb un total de 283 pacients.

Dos estudis varen comparar IDET i un PIRFT versus placebo (intervenció simulada). Un altre va comparar IDET amb un programa de rehabilitació (educació, exercicis i rehabilitació), un altre va comparar entre sí dues modalitats de PIRFT i un darrer va comparar IDET amb PIRFT.

La població d'estudi va consistir en pacients amb lumbàlgia crònica d'origen discogènic (establert per discografia de provocació o analgèsica) que havien fracassat a un tractament conservador previ per un temps molt variable.

El seguiment dels pacients va oscil·lar entre 8 setmanes i 6 mesos.

Tres estudis (AC) varen tenir una qualitat raonable. Els altres tres presentaven importants limitacions metodològiques que plantegen seriosos dubtes sobre el seu risc de biaix.

Un dels AC que va comparar IDET versus placebo va trobar una petita diferència significativa però clínicament irrellevant en el dolor i la discapacitat a favor de IDET. En canvi, a l'altre AC (el de major qualitat) no va haver-hi cap diferència, així com tampoc en la depressió, tolerància a la sedestació i a

l'esforç, consum de medicació i dèficits neurològics. L'AC que va comparar PIRFT amb placebo tampoc va trobar diferències en el dolor, discapacitat, qualitat de vida, efecte global, 'èxit' del tractament i consum de medicaments.

L'AC que va comparar entre sí dues modalitats de PIRFT no va trobar diferències entre elles. A més, va mostrar com en ambdós grups d'intervenció, la millora inicial observada sobre el dolor i la discapacitat havia desaparegut als 2 mesos (efecte fugaç).

Els dos estudis que varen obtenir uns resultats favorables a IDET quant a dolor i discapacitat varen consistir en estudis no aleatoritzats, amb un alt risc de biaix, la qual cosa qüestiona la seva validesa.

No es varen descriure efectes adversos a 2 AC amb IDET, mentre que a un altre es varen comunicar 4 casos de radiculopatia. L'AC que va comparar dues modalitats de PIRF va descriure un cas de discitis. Els estudis no aleatoritzats no van avaluar els efectes adversos.

En conclusió, l'evidència disponible no recolza l'eficàcia o efectivitat de les tècniques percutànies de termocoagulació intradiscal per al tractament de la lumbàlgia d'origen discogènic.

Publicació nº 3: *Kovacs FM, Urrútia G, Alarcón JD. Surgery versus conservative treatment for symptomatic lumbar spinal stenosis: a systematic review of randomized controlled trials. Spine (Phila Pa 1976). 2011;36:E1335-E1351.*

Factor d'impacte (2011): 2,510

Objectiu nº 3: Avaluar l'efectivitat de la cirurgia en comparació amb el tractament conservador sobre el dolor, discapacitat i qualitat de vida en l'estenosi lumbar simptomàtica.

La cerca actualitzada a juliol de 2009 va permetre identificar 5 AC amb un total de 918 pacients. Tots els estudis eren formalment d'alta qualitat, encara que la mida mostral d'un d'ells era molt petita i altres presenten algun problema que dificulta la interpretació de resultats.

La majoria de variables de resultat eren auto-reportades pels pacients. Degut a la naturalesa de les intervencions avaluades, no va ser possible l'emascament ni tampoc l'avaluació cegada de resultats, la qual cosa podria haver influït en alguna mesura en els resultats.

L'edat mitjana dels pacients oscil·lava entre 62 i 70 anys, tots ells amb estenosi espinal documentada per proves d'imatge. Un estudi va incloure només pacients amb espondilolistesi, un altre els va excloure, un altre només va incloure pacients amb espondilolistesi \leq grau I, i els altres dos varen incloure pacients amb i sense espondilolistesi. Un estudi va incloure pacients amb claudicació neurogènica, mentre que els altres quatre varen incloure pacients amb dolor irradiat a la cama, amb i sense claudicació (dades disponibles per a cada subgrup).

Quatre AC varen avaluar la cirurgia de descompressió (laminectomia, facetectomia medial, discectomia i/o eliminació d'osteofits i lligament groc

hipertròfic) amb o sense fusió (instrumentada/no instrumentada). El cinquè AC va avaluar in dispositiu interespinós de tracció (X-Stop), promogut per la indústria.

El tractament conservador del grup control va incloure una varietat de combinacions d'intervencions (ortosis, rehabilitació, teràpia física, exercicis, AINEs, analgèsics i educació, així com també calor i fred, TENS, ultrasons i injeccions epidurals d'esteroides), sempre a criteri del metge responsable del pacient.

El seguiment dels pacients va oscil·lar entre 2 i 10 anys (aquest darrer va consistir en una extensió).

Les pèrdues varen oscil·lar entre el 7,7% a 1 any i el 44% als 4 anys, encara que estaven equilibrades entre els grups quirúrgic i de tractament conservador.

La distància de claudicació va ser similar en ambdós grups. No obstant, tots els estudis consistentment varen mostrar una milloria significativa del dolor, la funció, la qualitat de vida i la satisfacció del pacient a favor de la cirurgia (amb o sense fusió). En general, la magnitud del benefici addicional produït per la cirurgia sobre el dolor i la funció va ser clínicament rellevant.

La cirurgia va ser més efectiva que el tractament conservador en els pacients amb i sense espondilolistesi, així com en els pacients amb o sense claudicació neurogènica, encara que la magnitud va ser major en els pacients amb claudicació.

Un problema habitual dels estudis va ser l'elevada taxa de creuaments entre els grups d'estudi (entre el 9% i 44% dels pacients), especialment cap al grup de cirurgia. No obstant, els dos AC amb una major taxa de creuaments presenten una anàlisi per ITT i una altra anàlisi alternativa segons el tractament rebut (*'as treated'*). Les diferències només varen ser significatives a favor de la cirurgia en l'anàlisi segons el tractament rebut, excepte la qualitat de vida (SF36 *bodily pain*) que va ser millor fins i tot a l'anàlisi per ITT.

L'avantatge observat amb la cirurgia es va posar de manifest prematurament (3-6 mesos) i es va mantenir constant fins als 2-4 anys, encara que a llarg termini les diferències entre els grups tendien a ser més petites.

El dispositiu interespinós de tracció també va obtenir uns millors resultats que el tractament conservador en el dolor, la funció, la qualitat de vida i la satisfacció dels pacients. Les diferències varen ser significatives a les 6 setmanes i es varen mantenir fins al final del seguiment als 2 anys. Els resultats varen ser favorables en els pacients amb i sense espondilolistesi (\leq grau I).

Entre els pacients que varen rebre cirurgia descompressiva, la taxa de complicacions durant la cirurgia va oscil·lar entre el 5,4% i 14% (la més freqüent, l'esquinç dural) i la taxa de complicacions post-operatòries entre el 8,2% i el 18%. La taxa de re-intervencions comunicada va ser del 1,3% al 2% a l'any, del 6% al 11% als dos anys i del 15% als 4 anys (dades d'un AC). La taxa de re-estenosi als 4 anys va ser del 5%. El nombre de morts va ser similar a ambdós grups, amb una taxa de mortalitat post-operatòria de 0,3% al 0,5%. Aproximadament el 90% dels pacients operats no varen patir cap episodi advers.

Quant al dispositiu interespinós, varen haver-hi poques complicacions derivades de la intervenció (n=8). La pèrdua sanguínia va ser només de 46 ml (molt inferior a la cirurgia de descompressió [198-589 ml]) i la durada del procediment de 54 minuts (molt inferior també a la cirurgia de descompressió [120-199 minuts]). El 96% dels pacients varen estar ingressats menys de 24 hores. El 7% dels pacients varen haver de ser intervinguts amb posterioritat per dolor persistent (laminectomia). Aquest percentatge va ser de l'11,9% en els pacients amb espondilolistesi.

En conclusió, en pacients amb estenosi espinal lumbar simptomàtica, la cirurgia descompressiva (amb o sense fusió) o la implantació d'un dispositiu interespinós de tracció són més efectius que el tractament conservador en pacients que han fracassat després de 3 a 6 mesos a tractament conservador.

Publicació nº 4: *Seco J, Kovacs FM, Urrútia G. The efficacy, safety, effectiveness, and cost/effectiveness of ultrasound and shock wave therapies for low back pain: A systematic review. Spine J. 2011 Apr 8. [Epub ahead of print]*

Factor d'impacte (2011): 3,024

Objectiu nº 4: Avaluar l'eficàcia, seguretat, efectivitat i cost-efectivitat de les teràpies basades en ultrasons i ones de xoc per la lumbàlgia.

La cerca actualitzada fins a desembre de 2008 va permetre identificar 4 AC, amb un total de 242 pacients, en general amb dolor i discapacitat moderades, amb o sense dolor irradiat.

Els ultrasons varen ser comparats amb placebo (ultrasons simulats), amb tracció lumbar, amb laser i amb manipulació espinal, i les ones de xoc amb TENS.

El seguiment dels pacients va oscil·lar entre 3 setmanes i 6 mesos.

Quant a la qualitat, dos estudis varen puntuar molt baix i varen ser considerats com d'alt risc de biaix. En general, les publicacions proporcionaven pocs detalls sobre els mètodes per poder fer una valoració acurada de la seva qualitat.

En un AC en pacients amb lumbàlgia aguda i dolor irradiat per hèrnia discal, la tracció lumbar, els ultrasons i el làser varen produir els mateixos resultats que els ultrasons quant a dolor i discapacitat a un i tres mesos.

Els altres tres AC es varen realitzar en pacients amb lumbàlgia crònica inespecífica sense dolor irradiat. En un AC (el de més baixa qualitat), el dolor va millorar aproximadament un 10% més amb ultrasons que amb placebo, però sense diferències en la mobilitat. En un altre, la manipulació espinal va millorar significativament el dolor als 6 mesos comparat amb ultrasons (un benefici de

1,5 punts en el VAS), així com també la discapacitat (5 punts a l'ODI), la flexió (9 mm al test de Schober) i l'extensió (3 mm).

En un AC, les ones de xoc varen tenir uns resultats similars que amb TENS quant a dolor, discapacitat, kinesiofòbia, 'eficàcia' percebuda, ansietat i depressió, i funció física.

Cap estudi va avaluar costos ni tampoc efectes adversos.

En conclusió, l'evidència disponible no recolza l'efectivitat dels ultrasons o les ones de xoc per al tractament de la lumbàlgia.

Publicació nº 5: Urrútia G, Bonfill X. Artículo especial. Análisis de la investigación española sobre dolencias del cuello y la espalda (1992–2006). Med Clin (Barc).2010;135(5):215–21

Factor d'impacte (2010): 1,413

Objectiu nº 5: Descriure la recerca espanyola sobre patologies mecàniques del raquis, el seu import, fonts de finançament i la seva productivitat.

La cerca electrònica i el contacte per correu (*mailing*) varen permetre identificar un total de 129 projectes d'investigació, 51 investigadors, 18 entitats finançadores i 80 publicacions sobre patologies mecàniques del raquis.

La taxa de resposta als requeriments d'informació per correu va ser baixa. No obstant, 13 dels investigadors que varen respondre eren responsables de 75 (58,1%) dels projectes identificats. Cap d'aquests 75 estudis va ser improductiu i a tots constava el finançament de forma transparent. Entre aquests, es trobaven totes les revisions sistemàtiques i 15 dels 16 assaigs clínics detectats. La resta d'investigadors (n=38) varen fer pocs estudis (la majoria un) o bé no varen respondre l'enquesta sobre finançament i publicacions.

Varen respondre l'enquesta per correu 79 serveis de 55 hospitals, d'un total de 1.100 serveis de 281 hospitals contactats (taxa de resposta del 7,2%). Disset (21,5%) havien participat en estudis de PMR. Això va permetre detectar projectes i publicacions addicionals a la cerca electrònica.

Varen respondre l'enquesta per correu 12 entitats finançadores, d'un total de 18 contactades. Només varen proporcionar dades dels projectes finançats. Els projectes denegats varen ser detectats mitjançant l'enquesta per correu a investigadors o als serveis

Quant al tipus d'investigació realitzada, van haver-hi 16 assaigs clínics, 9 revisions sistemàtiques i 98 d'altres tipus (fonamentalment estudis clínics) que

inclouen els 3 de recerca bàsica. Tots aquests tres estudis de recerca bàsica varen ser improductius pel que fa a publicacions.

Quant al finançament, les dues entitats que més projectes varen finançar va ser la Fundación Kovacs (FK) (n=51), una entitat privada sense ànim de lucre especialitzada en el mal d'esquena, i el ISCIII (n=24) a través del FIS (n=18) o l'AETS (n=6). La resta d'entitats varen finançar pocs estudis cadascuna, la majoria només un. Vuit estudis es varen fer sense finançament extern i en 40 (31%) no es disposava d'aquesta dada. En 6 projectes finançats no es va especificar la quantitat. Els cost total dels 75 projectes amb finançament transparent va ascendir a 13.136.490 d'euros (la FK va aportar el 90,5% d'aquest finançament i el ISCIII el 7%). D'aquests, 30 (40%) s'havien publicat.

Dels 51 projectes finançats per la FK, 19 (37%) havien estat publicats i 32 estaven en curs (globalment el 100%). Dels 18 finançats pel FIS, se n'havien publicat 6 (33%) i 5 més estaven en curs. Sis mes restaven sense publicar.

Dels 83 articles identificats, 8 corresponen a projectes sense finançament extern, 40 a projectes amb finançament desconegut i 32 a projectes amb finançament extern declarat. El seu factor d'impacte (FI) era molt variable, entre 0 i 18,13 (un article d'assaig clínic publicat a Lancet).

El FI total (FIT) dels projectes finançats i publicats per la FK va ser de 49,120, força més alt que els del FIS (7,793). El FI mig va ser de 2,729 i 1,113, respectivament. A més, dels 15 projectes rebutjats pel FIS, la FK en va finançar 12, dels quals 7 es van publicar amb un FIT de 16,190. El FI mig d'aquests projectes rebutjats (2,313) va ser superior al dels projectes que el FIS havia finançat (1,113).

Quatre entitats varen finançar els 12 projectes improductius detectats, dels quals 6 corresponen al FIS. Aquests estudis varen rebre un finançament de 560.350 euros, que correspon al 45,7% de la inversió total del FIS.

Neuroreflexotherapy for non-specific low-back pain (Review)

Urrútia G, Burton AK, Morral Fernández A, Bonfill Cosp X, Zanolli G



**THE COCHRANE
COLLABORATION®**

This is a reprint of a Cochrane review, prepared and maintained by The Cochrane Collaboration and published in *The Cochrane Library* 2009, Issue 4

<http://www.thecochranelibrary.com>



Neuroreflexotherapy for non-specific low-back pain (Review)
Copyright © 2009 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

TABLE OF CONTENTS

HEADER	1
ABSTRACT	1
PLAIN LANGUAGE SUMMARY	2
BACKGROUND	2
OBJECTIVES	2
METHODS	3
RESULTS	4
DISCUSSION	12
AUTHORS' CONCLUSIONS	13
ACKNOWLEDGEMENTS	13
REFERENCES	13
CHARACTERISTICS OF STUDIES	14
DATA AND ANALYSES	19
APPENDICES	19
WHAT'S NEW	19
HISTORY	19
CONTRIBUTIONS OF AUTHORS	20
DECLARATIONS OF INTEREST	20
NOTES	21
INDEX TERMS	21

[Intervention Review]

Neuroreflexotherapy for non-specific low-back pain

Gerard Urrútia¹, A Kim Burton², Antoni Morral Fernández³, Xavier Bonfill Cosp⁴, Gustavo Zanolí⁵

¹Centro Cochrane Iberoamericano. CIBER Epidemiología y Salud Pública (CIBERESP) Spain, Hospital de la Santa Creu i Sant Pau (U.A.B), Barcelona, Spain. ²Ergonomics and Clinical Biomechanics. Spine Research Unit, University of Huddersfield. UK, Huddersfield, UK. ³Escola Universitaria de Fisioteràpia, Universitat Ramon Llull, Santa Perpètua de la Mogoda (Barcelona), Spain. ⁴Iberoamerican Cochrane Centre. CIBER de Epidemiología y Salud Pública (CIBERESP), Spain. Hospital de la Santa Creu i Sant Pau, Universitat Autònoma de Barcelona, Barcelona, Spain. ⁵Orthopedics, Casa di Cura S. Maria Maddalena, Occhiobello (Ro), Italy

Contact address: Gerard Urrútia, Centro Cochrane Iberoamericano. CIBER Epidemiología y Salud Pública (CIBERESP) Spain, Hospital de la Santa Creu i Sant Pau (U.A.B), c/ Sant Antoni Maria Claret 171, Barcelona, 08041, Spain. gurrutia@santpau.cat.

Editorial group: Cochrane Back Group.

Publication status and date: New search for studies and content updated (no change to conclusions), published in Issue 4, 2009.

Review content assessed as up-to-date: 28 July 2009.

Citation: Urrútia G, Burton AK, Morral Fernández A, Bonfill Cosp X, Zanolí G. Neuroreflexotherapy for non-specific low-back pain. *Cochrane Database of Systematic Reviews* 2004, Issue 2. Art. No.: CD003009. DOI: 10.1002/14651858.CD003009.pub2.

Copyright © 2009 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

ABSTRACT

Background

Among the wide range of therapeutic alternatives proposed for the management of low-back pain (LBP), a less widely used technique from Spain, called neuroreflexotherapy (NRT) has claimed to show very favourable results, mainly in patients with chronic low-back pain.

Objectives

To assess the effectiveness of NRT for the treatment of non-specific LBP in adult patients, aged 16 to 65 years. A secondary objective was to compare NRT with other conventional interventions.

Search strategy

In July 2009, we updated the searches in CENTRAL (Issue 3, 2009), MEDLINE and EMBASE. No new trials were identified.

Selection criteria

Only randomised controlled trials (RCTs) of NRT for the treatment of patients with a clinical diagnosis of non-specific LBP were included.

Data collection and analysis

Two authors independently selected trials and extracted data using pre-designed forms. Because the outcome variables were not assessed in a homogenous way, it was not possible to pool the results to obtain an estimate of global effect, as initially planned.

Main results

Three RCTs were included, with a total of 125 subjects randomised to the control groups and 148 subjects receiving active NRT. Neuroreflexotherapy was the same in all three trials, while the control groups received sham-NRT in two trials and standard care in one. Two trials studied patients with chronic LBP, the third studied patients with a mix of chronic and sub-acute LBP. Clinical outcomes were measured in the short-term (15 to 60 days) in all three trials; in one trial, resource utilization was measured after one year. Individuals who received active NRT showed statistically significantly better outcomes than the control groups for measures of

pain, degree of mobility, disability, medication use, consumption of resources and costs. No significant differences were observed for quality of life measures. Side effects were more frequently reported in the control groups during short-term follow-up, with no major side effects reported by those receiving active NRT.

Authors' conclusions

NRT appears to be a safe and effective intervention for the treatment of chronic non-specific LBP. The efficacy is less clear for sub-acute LBP. However, these results are limited to three trials conducted by a small number of specifically trained and experienced clinicians, in a limited geographical location. No data are available on the ease and time-frame needed to achieve that level of expertise. RCTs by other practitioners, in other locations, that replicate the effects reported in this review are needed before recommending a broader practice.

PLAIN LANGUAGE SUMMARY

Neuroreflexotherapy for non-specific low-back pain

Neuroreflexotherapy, provided in specialized clinics in Spain, appears to reduce pain and disability for patients with chronic non-specific low-back pain.

Clinicians use a wide range of treatments to manage chronic low back pain. There is often little scientific support that they work, or that they would be useful for a broader population. In this review, neuroreflexotherapy performed better than placebo or standard care. However, until research duplicates these results in different settings, there is no strong evidence that it will work as well outside the specialty clinics in Spain.

BACKGROUND

A large variety of therapeutic interventions are available for the treatment of chronic low-back pain (LBP). However, the effectiveness of most of these interventions (educational, ergonomic or therapeutic) has not been convincingly demonstrated (Deyo 1996; Kaplansky 1998; van Tulder 1997a; van Tulder 2000) and consequently, the therapeutic management of chronic LBP varies widely.

Among the wide range of treatments proposed, a less widely used technique from Spain, called neuroreflexotherapy (NRT) has claimed very favourable results in the management of LBP. Neuroreflexotherapy is characterised by temporary implantation of a number of epidermal devices into trigger points in the back and into referred tender points in the ear. The theoretical physiological basis of this procedure has been outlined in depth in a number of publications (Kovacs 1993; Kovacs 1997; Kovacs 2002). Staples are implanted superficially into the skin, to a depth of less than two millimetres. No problems with pain or scarring associated with the procedure are reported in the literature. Neuroreflexotherapy is performed without anaesthesia and takes about 60 minutes for

the implantation. The staples remain in place for up to 90 days in the back and up to 20 days in the ear. Neuroreflexotherapy may be confused with acupuncture as both use puncture devices, but, according to the first author of the studies (personal communication), different zones of the skin are being stimulated. Acupuncture for LBP has been the subject of several systematic reviews, including a Cochrane review (Van Tulder 1999), but none have included NRT trials.

A large follow-up study in 2751 patients (Moreno 1992) and several clinical trials on NRT (Kovacs 1993; Kovacs 1997; Kovacs 2002) claiming surprisingly consistent, favourable results attracted our attention and stimulated our interest in a review of the possible effectiveness of this procedure.

OBJECTIVES

The aim of this review was to systematically assess the effectiveness of NRT for the treatment of non-specific LBP in adult patients, focusing on those of working age.

The principal objective of this review was to validly and reliably answer the questions:

- Is NRT effective for adult patients suffering from non-specific LBP?
- Is NRT safe in terms of the rate of adverse events, for the treatment of such patients?

The secondary objective was to compare NRT with other conservative interventions.

METHODS

Criteria for considering studies for this review

Types of studies

Randomised controlled trials (RCTs) and controlled clinical trials (CCTs), published in any language, were sought for this review. Trials were included by consensus after independent assessments by two authors (GU and AM), according to the following general criteria:

- a) RCTs and CCTs should have an appropriate control group: placebo for the principal objective and any other conventional intervention for the secondary objective.
- b) For the principal comparison, RCTs and CCTs should be blinded for patients and outcome assessors.

Types of participants

Adult subjects of working age (16 to 65 years old) with non-specific LBP that was not relieved with conservative treatment.

Patients with (sub)acute LBP (2 to 12 weeks) and/or chronic LBP (more than 12 weeks) were included.

All subjects with organically-caused back disorders (such as infection, metastatic diseases, neoplasm, osteoporosis, rheumatoid arthritis or fractures) were excluded, but imaging studies were not required to rule out organicity.

Types of interventions

Neuroreflexotherapy (defined as temporary implantation of epidermal devices into trigger points at the site of each subject's clinically involved dermatomes on the back and into referred tender points in the ear) was the therapy of interest. Acupuncture techniques were excluded from this review, since NRT and acupuncture are ostensibly (and probably actually) different procedures and there is already a Cochrane review on the latter (Van Tulder 1999).

For the primary objective, we considered placebo or sham treatment to be appropriate when the patients were given an intervention that did not penetrate the skin, or were given simulated treatment (the intervention was applied in purportedly non-reflexogenous areas or in reflexogenous areas that did not correspond with the involved lumbar dermatomes). Other conservative interventions were considered for comparisons for the secondary objective.

Types of outcome measures

RCTs and CCTs that used at least one of the five primary outcome measures that we considered to be the most important were included:

- a) Pain intensity.
- b) A global measure of well-being or quality of life.
- c) A global measure of disability for daily activities.
- d) Functional ability.
- e) Return to work (days off work).

Other secondary outcome measures for this review were:

- f) Medication use.
- g) Costs.
- h) Side effects.

These outcomes should have been measured and recorded before randomisation or before the start of the intervention (baseline) and again after the intervention period in order to assess change over time (up to three months for very short-term assessment, between three and six months for short/medium term assessment, and longer than six months for longer-term assessment).

Search methods for identification of studies

For the electronic search, we combined parts A and B of the Cochrane Back Group search strategy (van Tulder 1997) and our specific search strategy for NRT.

See Appendix 1 for the specific search for neuroreflexotherapy in MEDLINE (Ovid).

The specific search for neuroreflexotherapy was changed accordingly to search EMBASE (Ovid). See Appendix 2.

Relevant RCTs and CCTs meeting the inclusion criteria for this review were identified in the following steps:

(A) A computer-aided search of the MEDLINE (1966-2002) and EMBASE (1988-2002) databases using the search strategy recommended by the Editorial Board of the Cochrane Back Review Group (van Tulder 1997a).

(B) Screening of references in relevant reviews and in RCTs and CCTs identified in step A.

(C) Screening of the Cochrane Central Register of Controlled Trials, which is included in The Cochrane Library Issue 3, 2002.

(D) Personal communication with content experts in the field.

All articles were coded and details of source, intervention and population recorded. The author compiling the references (GU)

decided on potentially relevant trials and sent a full copy of them to the other authors (AM and KB) for inclusion assessment. In July 2009, we updated the searches in CENTRAL (Issue 3, 2009), MEDLINE and EMBASE. No new trials were identified.

Data collection and analysis

1. Study selection.

Two authors (GU and AM) searched The Cochrane Library, MEDLINE, and EMBASE and decided on potentially relevant trials. They independently reviewed potential articles and decided final eligibility according to the pre-determined selection criteria. One of the authors was a content-expert in the area (AM), while the other was not (GU). There was full agreement between them. Two other authors (KB and GZ) also read the full text of the articles, and concurred with the opinions of the first two authors. Due to the low number of papers and the easily recognisable characteristics of the technique, the assessment of eligibility criteria, data extraction and quality assessment could not be performed in a blinded fashion.

2. Methodological quality assessment.

The methodological quality of each article was independently assessed by three authors (GU, AM and KB) using the criteria list recommended by the Cochrane Back Review Group (van Tulder 1997a). Only the 10 items reflecting the internal validity of the RCTs were used. Equal weights were applied to all criteria. There was agreement among the three assessors.

3. Data extraction.

A specific form was designed and used by two authors (GU and AM) to independently extract relevant outcome data and descriptive information on the study population and the interventions from each trial. The third author (KB) checked the extraction, making minor corrections.

The first author of the three included trials was contacted by GU to obtain more detailed information on scarring of the skin after the procedure, the number of sub-acute patients included in the first trial and additional details on the NRT technique.

4. Data analysis.

For similar comparisons and outcome measures, we had planned to calculate an overall relative risk using the Mantel-Haenszel method for dichotomous outcomes, and weighted mean differences (or standardised mean differences if different instruments of measure were used) for continuous outcomes, using a fixed effect model or a random effects model (in case statistically significant heterogeneity was detected). We had also planned to perform a sensitivity analysis, using only the highest quality studies, in order to assess whether this dimension played a role in the results being observed. As the outcome measures used in the trials were clinically too heterogeneous, we performed a qualitative analysis, with a special emphasis on the methodological quality of the trials and the consistency of their findings. Sub-group analyses (sub-acute and chronic LBP) were not performed as data were not provided separately for these patients.

RESULTS

Description of studies

See: [Characteristics of included studies](#).

Only three RCTs, including 273 patients, fulfilling the inclusion criteria were electronically identified: [Kovacs 1993](#); [Kovacs 1997](#) and [Kovacs 2002](#). The characteristics of these studies are described in the *Table of Characteristics of Included Studies* and [Table 1](#); [Table 2](#) and [Table 3](#). All three trials were conducted by the same primary investigator, albeit with different research teams. [Kovacs 1993](#) and [Kovacs 1997](#) compared NRT with a placebo (sham intervention), which was given in the same way in both studies, matching the number of devices between experimental and control patients. The only difference between the two groups was the location of the epidermal devices, which in the control group were implanted within a 5 cm radius of the target zones. The burins fell out spontaneously (between 5 and 17 days) and the staples were left in place until they were removed at the end of the follow-up period (30 and 45 days in [Kovacs 1993](#) and [Kovacs 1997](#), respectively). [Kovacs 2002](#) compared NRT with standard care, described as counseling, drug treatment, possible laboratory tests, and imaging studies, potential referral to physiotherapy, rehabilitation for further treatment or to specialists for further evaluation.

Table 1. Results (Kovacs 1993)

Outcome	Description	Categories/ Measures	NRT	Sham	P
---------	-------------	-------------------------	-----	------	---

Table 1. Results (Kovacs 1993) (Continued)

Spontaneous pain relief	T30-T0 (self reported)	(self)	Disappeared + greatly improved / slightly improved + unchanged + worsened	45/2	1/41	<0.0001
Referred pain relief	T30-T0 (self reported)	(self)	Disappeared + greatly improved / slightly improved + unchanged + worsened	36/0	0/31	<0.0001
Tenderness pain relief	T30-T0 (self reported)	(self)	Disappeared + greatly improved / slightly improved + unchanged + worsened	46/1	0/42	<0.0001
Daily activity	T30-T0 (self reported)	(self)	Normal + greatly improved / slightly improved + unchanged + worse	36/5	1/36	<0.0001
Functional status	T30-T0 (physical exam)	(physical)	Normal + greatly improved / slightly improved + unchanged + worse	45/2	1/42	<0.0001
Changes in medication use	T30-T0 (self reported)	(self)	Suspended + reduced / maintained + increased	44/3	11/32	<0.0001
No. days off work	Self reported		Mean (SD)	4.70 (11.60)	12.50 (13.24)	<0.003
No. days in bed	Self reported		Mean (SD)	0.96 (4.92)	5.21 (9.50)	<0.008
Side effects	Spontaneously self-reported		Yes / No	4/43	17/26	<0.002

Table 2. Results (Kovacs 1997)

Outcome	Description	Categories/ Measures	NRT	Sham	P
Spontaneous pain relief	T45-T0 (VAS)	Mean difference (SD)	+3.09 (2.56)	+0.34 (2.98)	<0.001

Table 2. Results (Kovacs 1997) (Continued)

Referred pain relief	T45-T0 (VAS)		Mean (SD)	difference	+2.03 (2.49)	-0.61 (4.17)	0.003
Pain on movement relief	T45-T0 (VAS)		Mean (SD)	difference	+2.87 (3.01)	+0.03 (3.50)	<0.001
Anterior flexion	T45-T0 (VAS)		Mean (SD)	difference	+2.53 (3.07)	-0.09 (3.86)	0.033
Flexion to the right	T45-T0 (VAS)		Mean (SD)	difference	+2.28 (3.20)	-0.09 (4.16)	0.012
Flexion to the left	T45-T0 (VAS)		Mean (SD)	difference	+2.25 (2.79)	+0.14 (3.76)	0.012
Bending forward	T45-T0 (physical exam)		Mean (SD)	difference	-0.82 (10.56)	-5.38 (12.42)	0.096
Effect of LBP on quality of life: daily activities	T45-T0 (COOP chart)		Mean (SD)	difference	+0.81 (1.35)	+0.61 (1.38)	0.534
Effect of LBP on quality of life: social activities	T45-T0 (COOP chart)		Mean (SD)	difference	+0.26 (1.39)	+0.08 (1.50)	0.594
Effects of LBP on quality of life: pain during the past 6 weeks	T45-T0 (COOP chart)		Mean (SD)	difference	+1.13 (1.46)	+0.56 (1.18)	0.067
Effects of LBP on quality of life: change in condition	T45-T0 (COOP chart)		Mean (SD)	difference	+2.45 (1.11)	+2.83 (0.85)	0.095
Effects of LBP on quality of life: overall health	T45-T0 (COOP chart)		Mean (SD)	difference	+0.44 (0.89)	+0.25 (0.87)	0.340
Effects of LBP on quality of life: overall quality of life	T45-T0 (COOP chart)		Mean (SD)	difference	+0.16 (0.97)	+0.28 (0.85)	0.542
Physical condition	T45-T0 (COOP chart)		Mean (SD)	difference	+0.27 (1.26)	+0.44 (1.23)	0.164

Table 2. Results (Kovacs 1997) *(Continued)*

Side effects	Spontaneously self-reported	Yes/No	7/41	5/37	NS
--------------	-----------------------------	--------	------	------	----

Table 3. Results (Kovacs 2002)

Outcome	Description	Measures	Control	NRT	P
Improvement in LBP (VAS)	T60-T0 (self reported)	Median (range)	1.92 (-1.25, 3.04)	5.50 (3.73, 8.80)	0.000
Improvement in referred pain (VAS)	T60-T0 (self reported)	Median (range)	0.58 (-1.50, 2.01)	3.63 (2.69, 7.30)	0.001
Improvement in disability (Roland Morris Questionnaire)	T60-T0 (self reported)	Median (range)	2.05 (-1.50, 6.67)	8.67 (2.00, 13.33)	0.007
Improvement in quality of life (Euro-Quol)	T60-T0 (self reported)	Median (range)	-14.61 (-18.83, 22.50)	-11.67 (-50.00, 3.33)	0.628
Sick leave (days)	1 year (applicable for workforce people only)	Median (range)	105.2 (5, 330)	3.2 (0, 32.5)	0.001

In Kovacs 1993 and Kovacs 2002, patients were recruited from primary care consultations. In Kovacs 1997, they were recruited from the outpatient departments of three rheumatology services and one rehabilitation service of three hospitals. Patients with symptomatic spinal stenosis, clinically noticeable progressive motor weakness or sphincter impairment suggesting cauda equina syndrome were excluded. Kovacs 2002 used a cluster randomisation procedure, in which physicians (n = 21) rather than patients (n = 104) were selected as units of randomisation and analyses.

This review only considered the evaluation at the end of the follow-up periods to be of major clinical importance. This varied from 30 days (Kovacs 1993) to 45 days (Kovacs 1997) to 60 days (Kovacs 2002). This third trial also recorded resource utilization at one year. Since it was not feasible to blind the therapists, the outcome assessors were blinded to treatment allocation. In Kovacs 2002, the clinical outcome variables were evaluated by self-administered questionnaires (or telephone interviews) for the final outcome, whilst additional data were obtained by the family physicians, who were not blinded to the treatment allocation.

Risk of bias in included studies

The methodological quality of the three included trials is shown in detail in Table 4. There was no disagreement between the authors. A third author (KB) concurred with the opinions of the first two authors (GU and AM). The first two trials received a score of nine out of a possible ten, suggesting that appropriate measures had been put in place to minimise bias, unlike the last trial that only scored three out of a possible ten.

Table 4. Criteria for the Risk of Bias Assessment

Criteria for a judgment of yes for the sources of risk of bias
Method of randomisation adequate: A random (unpredictable assignment sequence. Examples of adequate methods are computer generated random number table and use of sealed opaque envelopes. Methods of allocation using date of birth, date of admission, hospital numbers, or alteration should not be regarded as appropriate.
Concealment of treatment allocation: Assignment generated by an independent person not responsible for determining the eligibility of the patients; this person has no information about the persons included in the trial and has no influence on the assignment sequence or on the decision about eligibility of the patient.
Blinding of patients: The reviewer determines if there was enough information about the blinding of the patient to score a yes.
Blinding of care providers: The reviewer determines if there was enough information about the blinding of the care provider to score a yes.
Blinding of outcome assessors: The reviewer determines if there was enough information about the blinding of the outcome assessors to score a yes.

Table 4. Criteria for the Risk of Bias Assessment (Continued)

Drop-out rate described and acceptable: The number of participants who were included in the study but did not complete the observation period or were not included in the analysis must be described and reasons given. If the percentage of drop-outs does not exceed 20% for short-term follow-up and 30% for long-term follow-up and does not lead to substantial bias, a yes is scored. (NB: these percentages are arbitrary, not supported by literature)
Similarity of baseline characteristics: In order to receive a yes, groups have to be similar at baseline regarding demographic factors, duration and severity of complaints and value of main outcome measure(s).
Co-interventions avoided or similar: Co-interventions should either be avoided in the trial design or comparable between the index and control groups.
Compliance acceptable: The reviewer determines if the compliance to the interventions is acceptable, based on the reported intensity, duration, number, and frequency of sessions for both the index intervention(s) and control intervention(s).
Timing outcome assessments similar: Timing of outcome assessment be identical for all intervention groups and for all important outcome assessments.

In [Kovacs 1993](#) and [Kovacs 1997](#), patients were allowed to continue their previously prescribed pharmacological treatment, and consumption of drugs was measured in both groups (type and amount) before the intervention and at the end of the follow-up period. In [Kovacs 1993](#), there were a few unexplained missing values, more in the control than in the treatment group. Only patients analysed at the end of the follow-up were considered for the analysis of each variable. If missing values were equated to therapeutic failures in the analysis, this would have favoured the experimental (NRT) group. In [Kovacs 1993](#), a Chi squared test was performed for all comparisons using categorical variables.

[Kovacs 2002](#) randomised physicians rather than patients and this randomisation was concealed. Neither patients nor care providers were blinded to the intervention and both groups of patients were free to continue with the local standard protocol for the management of LBP. The evolution was determined by the comparison of the value of the variables at the first and last assessments. Therefore, any missing value at any of the two assessments impeded such a comparison. There were missing values due to dropouts and exclusions (five patients), but there were some additional unexplained missing values in both groups. Although the paper only reports data from valid cases, results for the principal outcomes (improvement in pain and disability) were re-analysed by the authors of the original research, at our request, according to the intention to treat principle. The assumption made in this later analysis was that missing values were equivalent to no effect and were therefore assigned the same value as in the baseline.

Effects of interventions

Study Selection

Although a very limited number of trials on NRT were expected a priori because of the local use of this technique, we nevertheless conducted a full literature search. The Cochrane Library (Issue 3, 2002) identified 99 potential references, the MEDLINE search (1966-2002) identified 91, and the EMBASE search (1988-2002) 38.

After screening titles, abstracts and keywords, both authors (GU and AM) considered only three studies to be potentially eligible. Reviewing the full articles of these three studies, in collaboration with a third author (KB), confirmed their eligibility. None of the other identified studies evaluated NRT as defined for this review, although some used a similar term (reflex-therapy) referring to interventions such as manipulations, neural therapy and acupuncture.

[Kovacs 1993](#) included patients from primary care, with a minimum of seven days of LBP, however, the mean duration of pain was 84 weeks and 73 weeks for experimental and control groups respectively. [Kovacs 1997](#) included patients from rheumatology and rehabilitative services, with a minimum of twelve weeks of LBP -- the mean duration of LBP was 63 weeks in the control group and 86 in the experimental group. [Kovacs 2002](#) included patients from primary care, with a minimum of two weeks duration of LBP -- the mean duration of LBP was 48.13 days and 17.5 days for experimental and control groups respectively.

The literature search was updated in MEDLINE and EMBASE on November 5, 2004. No new trials on the effectiveness of NRT were identified, but there was a case report by Conde-Salazar that described a patient with a personal history of metal intolerance, who developed contact dermatitis, secondary to the insertion of the staples (Conde-Salazar 2004).

The literature search was updated in MEDLINE and EMBASE from 2004 to February, 2008 and in CENTRAL from 2002 to 2008 (Issue 1, 2008). The results were combined and the duplicates removed, leaving 324 unique references. There were no new reports of trials of the effectiveness of neuroreflexotherapy as a treatment intervention for LBP.

In July 2009, we updated the searches in CENTRAL (Issue 3, 2009), MEDLINE and EMBASE. No new trials were identified.

Effectiveness of Neuroreflexotherapy

A) Neuroreflexology versus sham Neuroreflexology.

Our search identified two RCTs comparing NRT to a sham intervention. It was not possible to conduct subgroup analyses for patients with sub-acute and chronic LBP because Kovacs 1993 only presented aggregated results.

Since none of the outcome variables were assessed in a homogeneous way, it was not possible to aggregate the results to obtain an estimate of global effect. Results of each outcome for each RCT are presented in Table 1; Table 2 and Table 3.

In this review, only the “residual benefit”, measured in Kovacs 1993, was considered to be of major clinical interest, with the “maximum beneficial effect” discarded as a transitory outcome. In Kovacs 1997, all patients underwent two separate evaluations at each follow-up assessments (the trial provides the score of each assessor but not an average of both assessments). As the concordance between these two assessors was very high, only the score obtained by the first of the two assessors was considered in this review for practical reasons.

A.1) Pain relief:

Both studies showed a statistically significant reduction in the severity of pain, whether local or referred, in the experimental group, at the end of the follow-up period (30 and 45 days, respectively). In Kovacs 1993, whereas almost all the patients in the NRT group (45 among 47) reported major relief or the disappearance of the pain, almost no-one in the control group (1 among 42) showed a similar improvement. In Kovacs 1997, the change in pain shows a large difference in favour of NRT. In this latter study, the observed size of the reduction in the follow-up assessment immediately after the intervention was 60% in the experimental group, but dropped to 50% when measured 45 days later. In the control group, an improvement of 20% was detected just after the

procedure and dropped to 10% 45 days later. Similar results were observed in the evaluation of “pain on pressure” (Kovacs 1993), “pain on movement” (forward flexion) and “pain experienced in the last six weeks” (Kovacs 1997).

A.2) A global measure of well-being or quality of life:

This measure was only assessed in Kovacs 1997, using the COOP chart. This is an instrument developed by The Dartmouth Primary Care Cooperative Information Project that assesses general dimensions of quality of life rather than the influence of any particular illness (Nelson 1987), and has been validated in Spanish. For the three dimensions related to perceived health or quality of life (from among a total of seven assessed with this instrument) a small favourable result was only detected for “Change in quality of life”. The differences found for the variables “overall health” and “overall quality of life” were not significant.

A.3) A global measure of disability for daily activities:

In Kovacs 1993, the observed improvement in the ability to perform daily activities, measured 30 days after the intervention using the COOP chart, was highly significant in the NRT group, with 88% of patients reporting a substantial improvement, whereas only three per cent of the control group showed this degree of improvement. In Kovacs 1997, there was no significant difference between the two groups in their ability to perform activities of daily living.

A.4) Functional ability:

In Kovacs 1993, patient’s “functional ability” at day 30 showed a significant improvement in the experimental group (96% of the patients), while only 2.3% of patients in the control group reported a significant improvement.

Kovacs 1997 evaluated the ability of the patients to bend forwards (fingertip-to-floor distance) as part of the physical examination (measurements were repeated three times and in each case the shortest distance was recorded). A greater and statistically significant improvement was seen in the experimental group at 45 days.

A.5) Return to work:

During the follow-up period, the mean numbers of days off work due to an episode of LBP in the experimental group in Kovacs 1993 was one third of that in the control group (4.7 versus 12.5 days) and statistically significant. The number of days off work were similar at baseline in both groups.

A.6) Medication use:

Both trials recorded medication consumption for pain, as reported by the patients. At the last assessment in [Kovacs 1993](#), 94% of patients in the experimental group reported that they had suspended or decreased the use of drugs, while this percentage was only 26% in the control group. This difference was not observed in [Kovacs 1997](#), where patient's consumption of drugs at baseline was sparse in both groups.

A.7) Side effects:

In [Kovacs 1993](#), patients were asked about adverse affects occurring during the follow-up period. These were more frequent in the control group (65% versus 9%), due mainly to gastric discomfort secondary to NSAID consumption, while local discomfort due to skin tautness caused by the devices in the hours following their implant, were reported equally (same percentage) in both groups. According to the original articles, none of these side effects were important enough to require the early extraction of the epidermal devices.

Pain involved in attaching the staples was not specifically reported in the trials, but seemingly no patient refused the intervention. No information is reported in the trials on scarring from the staples. According to the lead author who was contacted by the authors, skin scarring is not a relevant problem as staples are implanted very superficially in the skin and most scars disappear within one month.

A.8) Other outcomes:

The average number of days spent "lying in bed" in [Kovacs 1993](#) was one day for the intervention group and five days for the control group. This difference was statistically significant. There were no significant differences noted in [Kovacs 1997](#) for physical condition or social activities.

B) Neuroreflexology versus Standard Care.

[Kovacs 2002](#) examined a mixture of sub-acute and chronic patients, with pain lasting two or more weeks. All the outcomes related to effectiveness were assessed using specific scales at the initial visit and at two subsequent obligatory visits (at day 15 and 60). There was an additional follow-up assessment at day 354, when patients were interviewed by telephone regarding any treatment or diagnostic tests that had been prescribed since the last visit to the primary care centre. Data on duration of sick leave due to LBP were obtained from the register of the National Institute of the Social Security. For all clinical outcomes, the final ratings (day 60) were subtracted from those obtained in the first visit. Thus, effectiveness was determined by the variation of the median scores of the groups of patients who had received neuroreflexotherapy and those who had not. Between-group comparison were performed

using practices rather than patients as the units of analysis, using a non-parametric test. So, the outcome variables measuring effectiveness were averaged over all patients per physician.

B.1) Pain relief:

A statistically significant reduction in the severity of pain was reported in the experimental groups at the end of the follow-up period (at 60 days), although there was an improvement in both experimental and control groups (median: NRT 5.50 [range 3.73 to 8.80] vs Control 1.92 [range -1.25 to 3.04]; $p < 0,000$). For referred pain, the medians were NRT 3.63 [range 2.69 to 7.30] vs. Control 0.58 [range -1.50 to 2.01]; $p < 0.001$).

B.2) Disability:

There was a statistically significant improvement in disability at the end of the follow-up period (at 60 days) reported by those who had received NRT (median: NRT 8.67 [range 2.00 to 13.33] vs Control 2.05 [range -1.50 to 6.67]; $p < 0.007$).

B.3) Quality of life:

The trial measured quality of life by the self-administered EuroQol questionnaire (using the validated Spanish version). Although all groups tended to report some improvement in their quality of life, there was no statistically significant difference at the end of the follow-up period (at 60 days) between those who had received NRT and those who had not.

B.4) Sick leave:

Information regarding the number of days on sick leave during the one-year follow-up period was obtained for all patients from the register of the National Institute of the Social Security. Patients in the experimental groups experienced a statistically significantly shorter duration of sick leave (median 3.2 days; range 0 to 32.5) compared to the control groups (median 105.2 days; range 5 to 330).

B.5) Consumption of resources during follow-up:

During the one-year follow-up period, participating physicians recorded the use of health care services resulting from the management of LBP. Patients in the experimental groups showed statistically significantly fewer visits to private and public specialists, fewer radiographic studies prescribed in the primary care setting, and a lower cost of drug treatment.

B.6) Side effects:

Patients were asked about adverse affects occurring during the follow up period. Two patients in the control groups reported epigastralgia and treatment with nonsteroidal anti-inflammatory drugs

had to be withdrawn. Side effects related to NRT (such as transient tightness, skin irritation, and infection of the skin secondary to implantation of the surgical material) were not reported during the trial.

B.7) Cost-effectiveness:

The trial included a cost-effectiveness analysis in which the cost per point of improvement on the corresponding scale was used as a measure of efficiency. They conducted sensitivity analyses according to three specific assumptions: a) the most optimistic, b) the most conservative and c) the average assumption. In all cases, cost-effectiveness ratios for pain and disability were reported to be favourable for the experimental group.

DISCUSSION

Neuroreflexotherapy is unusual in that it seems to be currently performed in a limited geographical region (Spain), by a relatively small number of practitioners. Furthermore, the same principal investigator (who is also a leading NRT practitioner) was involved in all of the published RCTs. Nevertheless, those facts are not an a priori barrier to a review of the effectiveness of the therapy, if suitable published trials are available for analysis.

Overall Results

A) Neuroreflexology versus sham Neuroreflexology.

The reported results from the two published randomised trials comparing NRT with sham-NRT show a statistically significant short-term positive effect on chronic back pain for the main outcomes of pain, ability to perform daily activities, and functional ability, as well as secondary outcomes of return to work, side effects and medication use, when the treatment was given by a very limited number of particularly experienced physicians (a total number of two in these trials). Although the follow-up period in both trials (30 and 45 days) is not sufficient to assess the effect of NRT on recurrence of back pain, it can be considered adequate for assessing the effect of the intervention during the current episode. The effect appeared to be rapid and remained for at least six weeks after intervention in most of patients treated.

In spite of differences in the reduction of pain and disability between the treated group and the control group, there were no substantial differences reported in their quality of life or perceived health as measured by instruments which are not specific for low-back pain.

The extent of the reported clinical advantage of NRT over sham treatment is unusually high compared with other trials of treatment for back pain. A surprising lack of improvement was seen in

any of the outcome measures for the patients in the control groups. A possible explanation is that the instrument used in Kovacs 1993 (categorical outcomes) was sensitive to the level of change in the intervention group but not the seemingly smaller changes in the control group. In Kovacs 1997, the data tables in the original study do show improvements in the control group, though these are very small and much less than the intervention group. Bearing in mind the short follow-up and the fact that these were chronic pain patients, it is conceivable that this could represent the 'placebo' response for this type of patient.

A limitation of these trials is that one might normally expect a longer follow-up for what is a somewhat invasive procedure, but it is the authors' impression from the published papers that NRT is not an aggressive procedure likely to have long-term detrimental consequences.

Blinding of patients is a crucial issue in trials on LBP interventions. Kovacs 1993 did not provide detailed information on this. Nevertheless, the information given in both original papers suggests that intervention and control groups were handled in the same fashion, and it is said that the number and general location of staples/burins was largely matched between experimental and control groups. So, taking into account the high level of (reported) similarity between real and sham NRT, it is doubtful that the patients could detect real from sham unless they had specialist knowledge of the NRT procedure.

B) Neuroreflexology versus Standard Care.

The results from one recent randomised trial of NRT as a supplement to standard management protocol for LBP in routine general practice show a statistically significant short-term (60-days) effect on pain relief (local and referred) and ability to perform daily activities, and for duration of sick leave and consumption of resources throughout the one-year follow-up period. Again, in spite of differences in the reduction of pain and disability between the treated and control groups, there was no difference reported in their quality of life as measured by instruments that are not specific for low-back pain.

There is some concern about the reliability of the one-year assessment of consumption of resources by means of a telephone call, as recall bias may have occurred. Nevertheless, the difference between groups, as well as the sensitivity analysis performed by the authors, suggests that NRT reduces to some extent the cost of the management of LBP compared with standard care.

The imbalances observed between groups at baseline is a potential problem with cluster randomisation. Nevertheless, these differences do not invalidate the general conclusions of the trial, as they suggest a worse prognosis in patients in the intervention group, which is against the hypothesis demonstrated in the study.

The extent of the reported changes in the short-term (60 days) for clinical outcomes, and reduced sick leave and lower health-care costs over a 12-month period, suggest that NRT can be an effective

and safe supplement to the standard primary care management protocol for LBP in Spain.

POTENTIAL FOR SELECTION BIAS:

We identified only three published RCTs of NRT, all of them completed in Spain. However, this is a technique that is (confidently) believed to be used in a limited geographical area only, so we believe our search strategy had a low potential for study-selection bias.

METHODOLOGICAL QUALITY:

The methodological quality of the RCTs included in this review was reasonably high. All of the RCTs seemed to satisfactorily meet the essential methodological criteria related to randomisation, concealment of allocation, and blinding of the response assessment; we did not identify obvious reasons to suspect bias with regard to these issues. With respect to losses and dropouts, there are some doubts, because not all patients randomised were accounted for in the analysis of each variable. However, in view of the apparent effect size, imputing missing values as failures would be unlikely to affect the conclusions in this instance.

AUTHORS' CONCLUSIONS

Implications for practice

The main finding of this review is that NRT appears to be a safe and effective intervention for the short-term treatment of chronic non-specific LBP. That said, it is important to add some caveats.

These results are currently limited to trials conducted only in one country, by a small number of specially trained practitioners. This makes it difficult to recommend the wider use of NRT without first conducting randomised controlled trials by other practitioners, in other countries.

There are no objective reasons to suppose that socio-demographic, biological or cultural variables could affect the overall effects on pain and functional status. However, psychosocial factors may influence the degree of disability and the rate of return to work, and therefore results on these variables may vary from one setting to another.

Implications for research

Generalisation of the results remains difficult to establish, partly because of the very limited number of clinicians involved in the published trials, and partly because the results may only be valid when highly trained physicians are performing the interventions. No data are available on the ease and time-frame needed to achieve that level of expertise. Therefore, if further RCTs are undertaken in other settings, it is fundamental to assess the degree of competence that therapists would need to reach in order to achieve similar results to those observed in these trials.

Effectiveness of NRT in patients with (sub)acute LBP has not been clearly demonstrated. A well designed RCT with larger samples and precise patient selection criteria is needed to answer this question.

It needs to be determined if the unusually positive results in the current trials can be replicated. These trials would ideally, but not necessarily, be multi-centered and multinational, with the involvement of different clinical teams not previously linked to the authors of the original studies. That would also allow a better estimation of both the efficacy and effectiveness of NRT.

ACKNOWLEDGEMENTS

We would like to thank Dr. Maurits van Tulder for his advice while writing the protocol for the review. We also want to thank the first author of the three existing trials on NRT (Dr. FM Kovacs) for providing more detailed information than that previously published.

REFERENCES

References to studies included in this review

Kovacs 1993 *{published data only}*

Kovacs FM, Abraira V, López-Abente G, Pozo F. Neuro-reflexotherapy intervention in the treatment of nonspecific low back pain: a randomized, double-blind, controlled clinical trial [La intervención neuroreflejoteràpica en el tratamiento de la lumbalgia inespecífica: un ensayo clínico controlado, aleatorizado, a doble ciego]. *Med Clin (Barc)* 1993;**101**:570–5.

Kovacs 1997 *{published data only}*

Kovacs FM, Abraira V, Pozo F, Kleinbaum DG, Beltrán J, Mateo I, et al. Local and remote sustained trigger point therapy for exacerbations of chronic low back pain [Local and remote sustained trigger point therapy for exacerbations of chronic low back pain]. *Spine* 1997;**22**:786–97.

Kovacs 2002 *{published data only}*

Kovacs FM, Llobera J, Abraira V, Lázaro P, Pozo F, Kleinbaum D. Effectiveness and cost-effectiveness analysis of neuroreflexotherapy

for subacute and chronic low back pain in routine general practice. *Spine* 2002;**27**(11):1149–59.

Additional references

Conde-Salazar 2004

Conde-Salazar L, Valks R, Malfeito JE, Garcia C, Ledo S. Contact dermatitis from the staples of neuroreflexotherapy. *Contact Dermatitis* 2004;**51**(4):217–8.

Deyo 1996

Deyo RA. Drug therapy for back pain. Which drugs help which patients?. *Spine* 1996;**21**:2840–2849.

Kaplansky 1998

Kaplansky BD, Wei FY, Reecer MV. Prevention strategies for occupational low back pain. *Occupational Medicine* 1998;**13**:33–45.

Moreno 1992

Moreno J, Gestoso M, Kovacs FM. Effectiveness of the neuro-reflexotherapeutic intervention in the treatment of mechanical spinal disorders: preliminary results [La efectividad de la intervención neurorreflejoteràpica en el tratamiento de la patología mecánica del raquis: resultados preliminares]. *Med Trab* 1992;**1**:433–443.

Nelson 1987

Nelson E, Wasson J, Kirk J, Keller A, Clark D, Dietrich A, Steward A, Zubkoff M. Assessment of function in routine clinical practice: description of the COOP Chart method and preliminary findings. *J Chronic Dis*. 1987; Vol. 40, issue Suppl 1:55S–69S.

van Tulder 1997

van Tulder MW, Assendelft WJJ, Koes BW, Bouter LM, the Editorial Board of the Cochrane Collaboration Back Review Group. Method guidelines for systematic reviews in the Cochrane Collaboration Back Review Group for Spinal Disorders. *Spine* 1997;**22**:2323–2330.

van Tulder 1997a

van Tulder MW, Koes BW, Bouter LM. Conservative treatment of acute and chronic non-specific low back pain: a systematic review of randomized controlled trials of the most common interventions. *Spine* 1997;**18**:2128–56.

Van Tulder 1999

van Tulder MW, Cherkin DC, Berman B, Lao L, Koes BW. Acupuncture for low back pain (Cochrane review). *Cochrane Database of Systematic Reviews* 2004, Issue 1. [DOI: 10.1002/14651858.CD001351.pub2]

van Tulder 2000

van Tulder MW, Esmail R, Bombardier C, Koes BW. Back schools for non-specific low back pain (Cochrane Review). *Cochrane Database of Systematic Reviews* 2004, Issue 1. [DOI: 10.1002/14651858.CD000261.pub2]

References to other published versions of this review

Urrútia 2004

Urrútia G, Burton AK, Morral A, Bonfill X, Zanolli G. Neuroreflexotherapy for non-specific low-back pain. *Cochrane Database of Systematic Reviews* 2004, Issue 2. [DOI: 10.1002/14651858.CD003009.pub2.]

* Indicates the major publication for the study

CHARACTERISTICS OF STUDIES

Characteristics of included studies [ordered by study ID]

Kovacs 1993

Methods	Allocation to the study groups was established according to a table of random permutations. Sealed opaque envelopes with correlative Arabic numerals on the front contained the corresponding number of the table. Both the outcome assessor and patients were unaware of the group where patients had been assigned.
Participants	91 adult patients (20 to 65 years old) presenting an episode of LBP refractory to conventional medical treatment lasting for more than 7 days consecutively recruited from primary care consultations in Palma de Mallorca (Spain). Both groups were fully comparable for all the main variables. Mean duration of the current episode = 18 months. Mean duration of LBP syndrome for all the participants = 8 years.
Interventions	Patients in the treatment group received a single NRT intervention. Patients in the control group underwent a similar procedure although inappropriate zones were stimulated (sham intervention). Two therapists treated patients. Patients in both groups were allowed to continue drug and physiotherapy treatments as prescribed by their GPs.
Outcomes	1) Measures of change with respect the baseline (Day 30-Day 0): a) Pain relief, b) Daily activity, c) Medication use, d) Degree of mobility. 2) Others as number of days off work, number of days laying in bed, side effects.
Notes	Results at the baseline (before the intervention), immediately after intervention (at 15 minutes) and at short term (30 days) are provided.

Risk of bias

Item	Authors' judgement	Description
Adequate sequence generation?	Yes	
Allocation concealment?	Yes	A - Adequate
Blinding? All outcomes - patients?	Yes	
Blinding? All outcomes - providers?	No	
Blinding? All outcomes - outcome assessors?	Yes	

Kovacs 1993 (Continued)

Incomplete outcome data addressed? All outcomes - drop-outs?	Yes	drop-outs - less than 20%
Similarity of baseline characteristics?	Yes	
Co-interventions avoided or similar?	Yes	patients were allowed to continue pre-trial meds if needed
Compliance acceptable?	Yes	
Timing outcome assessments similar?	Yes	

Kovacs 1997

Methods	Randomization was carried out according to a table of random permutations. Only the person responsible for randomising patients and the therapist knew the result of the assignment. Both the outcome assessors and patients were unaware of the group where patients had been assigned.
Participants	78 adult patients (27 to 65 years old) presenting an episode of LBP refractory to conventional treatment lasting for more than 12 weeks consecutively recruited from outpatient clinics at the hospital level in Madrid (Spain). Both groups were fully comparable for all the main variables. Mean duration of the current episode = 63 and 86 weeks. Mean duration of LBP syndrome = 8,9 and 9,2 years.
Interventions	Patients in the treatment group received a single NRT intervention. Patients in the control group underwent a similar procedure although inappropriate zones were stimulated (sham intervention). The same therapist treated all patients. Patients in both groups were allowed to continue drug and physiotherapy treatments as prescribed by their GPs.
Outcomes	1) Measures of change with respect the baseline (Day 45-Day 0): a) Pain relief, b) Daily activity, c) Medication use, d) Degree of mobility. 2) Others as number of days off work, number of days laying in bed, side effects.
Notes	Results at the baseline (before the intervention), immediately after intervention (at 5 minutes) and at short term (45 days) are provided.

Risk of bias

Item	Authors' judgement	Description
Adequate sequence generation?	Yes	
Allocation concealment?	Yes	A - Adequate

Kovacs 1997 (Continued)

Blinding? All outcomes - patients?	Yes	
Blinding? All outcomes - providers?	No	
Blinding? All outcomes - outcome assessors?	Yes	
Incomplete outcome data addressed? All outcomes - drop-outs?	Yes	less than 20%
Similarity of baseline characteristics?	Yes	
Co-interventions avoided or similar?	Yes	patients were allowed to continue pre-trial meds if needed
Compliance acceptable?	Yes	
Timing outcome assessments similar?	Yes	

Kovacs 2002

Methods	Cluster randomisation (21 voluntary GPs were randomised). Each doctor recruited a median of 2.5 (Control) and 3 patients (NRT group) (range 1 to 14).
Participants	104 adult patients (28 to 61) presenting an episode of LBP lasting at least 14 days in spite of conventional treatment were consecutively recruited from primary care consultations in Palma de Mallorca (Spain). Intensity of pain and duration of current episode were higher in the NRT group while length of time on sick leave before inclusion was slightly higher in the control group. Median duration of the current episode of LBP = 17.5 days (Control) and 48.13 (NRT group). About 90% in the NRT group and 82% in the control group had experienced one or more previous episodes.
Interventions	Patients in the treatment group received NRT intervention (mean number of procedures 1.44) in addition to the standard care for LBP in the primary care setting. Patients in the control group received the so-called standard protocol for LBP.
Outcomes	1) Measures of change with respect the baseline (Day 60-Day 0): a) Pain relief (local and referred), b) Disability, c) Quality of life, and d) Side effects. 2) Measures at the end of 1-year follow-up period: a) number of days off work, b) consumption of resources.
Notes	
<i>Risk of bias</i>	

Kovacs 2002 (Continued)

Item	Authors' judgement	Description
Adequate sequence generation?	Yes	
Allocation concealment?	No	As physicians in this study were randomized not to treat with A versus B, but to refer patients or not to refer patients to a specialized unit on Neuroreflexotherapy, we believe that in this particular case the risk of selection bias was low.
Blinding? All outcomes - patients?	No	not applicable
Blinding? All outcomes - providers?	No	not applicable
Blinding? All outcomes - outcome assessors?	No	not applicable
Incomplete outcome data addressed? All outcomes - drop-outs?	No	
Similarity of baseline characteristics?	No	
Co-interventions avoided or similar?	Yes	Patients were allowed to continue pretrial meds if needed
Compliance acceptable?	Unclear	Unclear from text
Timing outcome assessments similar?	Yes	

DATA AND ANALYSES

This review has no analyses.

APPENDICES

Appendix 1. MEDLINE search strategy

- 1.SURGICAL-STAPLING\$.SH
- 2.SURGICAL-STAPLERS\$. SH
- 3.PROSTHESIS-IMPLANTATION. SH
- 4.PROSTHESES-AND-IMPLANTS.SH
- 5.IMPLANTS-EXPERIMENTAL\$.SH
- 6.SKIN.SH
- 7.EPIDERMIS.SH
- 8.EAR\$.SH
- 9.REFLEXOTHERAPY\$.SH
- 10.PHYSICAL-STIMULATION.SH
- 11.NEUROREFLEXOTHERAP\$
- 12.NEURO-REFLEXOTHERAP\$
- 13.or/1-12

Appendix 2. EMBASE search strategy

1. NEUROREFLEXOTHERAP*
2. explode "NERVE-STIMULATION"/all subheadings
3. #1 OR #2

WHAT'S NEW

Last assessed as up-to-date: 28 July 2009.

29 July 2009	New search has been performed	literature search updated. no new RCTs identified.
--------------	-------------------------------	--

HISTORY

Protocol first published: Issue 2, 2001

Review first published: Issue 2, 2004

3 June 2008	Amended	Converted to new review format.
1 February 2008	New search has been performed	In February 2008, we updated the searches in CENTRAL (Issue 1, 2008), MEDLINE and EMBASE. No new trials were identified.
1 November 2004	New search has been performed	The literature search was updated in MEDLINE and EMBASE in November 2004. No new trials on the effectiveness of NRT were identified, but there was a case report by Conde-Salazar that described a case of contact dermatitis in an individual with a personal history of metal intolerance, secondary to the insertion of the staples. This information has been added to the review.

CONTRIBUTIONS OF AUTHORS

GU and AM independently applied the selection criteria, assessed the quality of included trials and extracted the data.

KB and GZ independently read the trials and gave advice on the opinion of the two reviewers regarding the selection criteria, quality assessment and interpretation of results.

GU and KB wrote the draft of the review.

XB gave external advice in the process of the review.

DECLARATIONS OF INTEREST

None of the authors have any experience in performing NRT, have not participated in any of the trials included in this review, nor have any kind of professional involvement with the investigators or clinics in the trials reviewed. The authors undertook the task for its scientific and clinical interest only. The author who initially registered the title (GU), although not working in this field, chose this topic mainly because of the impact of this procedure on the Spanish social mass media. Since then, GU has kept relatively close contact with Dr. Kovacs in relation to this and other initiatives (a grant application). One author (GZ) visited one NRT clinic in Palma de Mallorca in April 2002, whilst another (KB) has worked with the lead author of the trials on an unrelated initiative.

NOTES

Note from the Co-Editors. This review addresses all of the following comments in their discussion, however, the Co-editors wish to echo the unusual nature of the evidence available to date . The effect of Neuroreflexotherapy is studied in three RCTs with 273 patients with low-back pain. Dr. Kovacs was the principal investigator for all three trials, albeit with different research teams. Patients in the trials were recruited from primary care settings and treated in three Kovacs Foundation clinics (A Spanish private, non-profit research and medical institution). This association does not imply methodologically flawed trials. In fact, the review was clear that the trials complied with the Back Group's methodological standards.

However, we would feel more reassured if similar evidence was available from RCTs conducted in other countries, with other care providers and different researchers. The very large positive response in the intervention group compared to the placebo group is unusual for trials in chronic back pain. Duplication of these results in other settings would also enable better assessment of the generalizability of these findings.

Therefore, we concur that until research duplicates these results in different settings, there is no strong evidence that it will work as well outside the specialty clinics in Spain.

INDEX TERMS

Medical Subject Headings (MeSH)

Acute Disease; Chronic Disease; Low Back Pain [*therapy]; Randomized Controlled Trials as Topic; Reflexotherapy [*methods]; Transcutaneous Electric Nerve Stimulation [*methods]

MeSH check words

Humans

Percutaneous Thermocoagulation Intradiscal Techniques for Discogenic Low Back Pain

Gerard Urrútia, MD,*† Francisco Kovacs, MD, PhD,†‡ Maria Betina Nishishinya, MD,*† and Javier Olabe, MD, PhD§†

Study Design. Systematic review.

Objective. To systematically review the evidence on the efficacy, effectiveness, and safety of percutaneous thermocoagulation intradiscal techniques for discogenic low back pain.

Summary of Background Data. The intervertebral disc is thought to be the source of pain in a relevant proportion of cases of low back pain (LBP). Two percutaneous thermocoagulation intradiscal techniques have been described to treat discogenic LBP: percutaneous intradiscal radiofrequency thermocoagulation (PIRFT) and intradiscal electrothermal therapy (IDET).

Methods. An electronic search was performed in MEDLINE, EMBASE, and the Cochrane Library databases up to 2005, to identify nonrandomized controlled trials and randomized controlled trials (RCTs) on those techniques. All relevant studies were methodologically assessed independently by 3 reviewers. RCTs were assessed following the criteria recommended by the Cochrane Back Review Group. A qualitative synthesis of results was performed.

Results. Six studies were included with a total of 283 patients. Two open, nonrandomized trials (95 patients) showed positive results for IDET compared with rehabilitation and PIRFT. Results from 2 RCTs showed no differences between PIRFT and placebo, and between different PIRFT techniques. Two RCTs compared IDET with placebo. One suggested differences only in pain and in disability, while the best quality RCT showed no differences.

Conclusions. The available evidence does not support the efficacy or effectiveness of percutaneous thermocoagulation intradiscal techniques for the treatment of discogenic low back pain.

Key words: intradiscal electrothermal therapy, percutaneous intradiscal radiofrequency thermocoagulation, low back pain, systematic review, clinical trial. **Spine 2007; 32:1146-1154**

Low back pain (LBP) is one of the most common musculoskeletal conditions. Although in most cases it is impossible to identify the cause of pain, it has been proposed that the disc could be a source of pain in a major proportion of patients. Discogenic LBP might be due to any cause stimulating the sensory nerve endings of the disc.¹⁻⁵ It has been proposed that discogenic LBP should be suspected when seeing a patient with nonradicular pain in the absence of neural tension signs, spinal deformity, and instability,⁶ with a positive provocative discography test⁷⁻¹¹ or a $\geq 50\%$ pain improvement after an anesthetic disc injection.¹² However, the validity of those techniques to identify patients with discogenic pain is uncertain.¹³ The natural history of that disorder is actually not known; its pathogenesis remains unclear and treatment mainly aims at reducing pain and improving symptoms.¹²

During the last decades, minimally invasive techniques for discogenic LBP have been developed as alternative to conventional surgery, in order to try to improve clinical results and decrease adverse events, risks, and costs. Among those, percutaneous thermocoagulation intradiscal techniques are based on the coagulation of the disc caused by heat.

Initially developed by Saal and Saal,¹⁴ percutaneous thermocoagulation intradiscal techniques are a minimally invasive treatment option for discogenic LBP. They consist on the insertion and heating of an electrothermal catheter into the disc under fluoroscopic guidance, with local anesthesia and light intravenous sedation. There are 2 different techniques, depending on whether the heat is generated by radiofrequency (percutaneous intradiscal radiofrequency thermocoagulation [PIRFT]) or electrically (intradiscal electrothermal annuloplasty, or intradiscal electrothermal therapy [IDET]). With IDET, once the catheter is in place into the posterior anular wall of the affected disc, its end is heated to 65°. If the patient tolerates the pain 1 minute, the temperature is increased 1°C every 30 seconds until reaching a temperature from 80° to 90°C. The process takes around 15 minutes and, once finished, the catheter is removed and an antibiotic is injected into the disc to prevent infection. With PIRFT, the catheter is placed into the center of the disc rather than the anulus. Then, the

From *Centro Cochrane Iberoamericano, Servei d'Epidemiologia i Salut Pública, Hospital de la Santa Creu i Sant Pau, Universitat Autònoma de Barcelona, Barcelona, Spain and Red Temàtica de Medicina Basada en la Evidència; †Departamento científico, Fundación Kovacs, Palma de Mallorca, Spain; ‡Red Española de Investigadores en Dolor de la Espalda; and §Servicio de Neurocirugía, Clínica Juaneda, Palma de Mallorca, Spain.

Acknowledgment date: April 10, 2006. First revision date: July 10, 2006. Second revision date: August 30, 2006. Acceptance date: September 29, 2006.

Supported by a grant from the Instituto de Salud Carlos III (Agency for Health Technology Assessment, Grant No. 03/10003), Fundación Kovacs, Centro Cochrane Iberoamericano.

The device(s)/drug(s) is/are FDA-approved or approved by corresponding national agency for this indication.

Institutional and Foundation funds were received in support of this work. No benefits in any form have been or will be received from a commercial party related directly or indirectly to the subject of this manuscript.

Address correspondence and reprint requests to Gerard Urrútia, Centro Cochrane Iberoamericano, Hospital de la Santa Creu i Sant Pau, Sant Antoni M. Claret 171, 08041 Barcelona, Spain; E-mail: gurrutia@santpau.es

device is activated for 90 seconds at a temperature of 70°C. The objective of both techniques is to produce a retraction on the collagen fibrils, and the thermocoagulation of inflammatory tissue and the nervous fibrils of the disc.

Results from case series have suggested promising results for PIRFT and IDET, but no comprehensive systematic review of controlled trials, including randomized controlled trials, has been performed so far. The aim of this study was to systematically review the efficacy, effectiveness and safety of IDET for the treatment of nonradicular LBP in the absence of major spinal structural changes, in which the disc was suspected to be the cause of pain (hereafter referred to as “discogenic” pain).

Materials and Methods

Search and Study Selection. An electronic search was performed in the Cochrane Library 2005 (Issue 4), MEDLINE (1966–2005), and EMBASE (1980–2005). To maximize the sensitivity of the search, no methodologic filters to identify randomized controlled trials were used, and only terms related to the technologies to be assessed were used. The search strategy is shown in the appendix.

Studies were included in this review if they were controlled (prospective nonrandomized with intrastudy comparison) or randomized controlled trials (RCTs), they assessed IDET or PIRFT in patients with LBP without radiculopathy or major structural changes, in which provocative or analgesic disc injections were used to try to establish the diagnosis of “discogenic” LBP, and assessed at least one of the following variables: pain severity, disability, quality of life, daily activities, impact in work, patient satisfaction, medication use, side effects, health resources employed, total costs, rate of early reintervention, and rate of late reintervention.

Quality Assessment and Data Analysis. Methodologic quality was independently assessed (by M.B.N., G.U., and F.K.) following the criteria recommended by the Cochrane Back Review Group.¹⁵ Data extraction was independently undertaken by (by M.B.N., G.U., and F.K.) and any differences were resolved by consensus. A quantitative synthesis of data was not possible because of heterogeneity of the methods used across the studies. Therefore, a qualitative analysis of data was performed, based on the methodologic quality of included trials and the consistency of their findings.

Results

Study Characteristics

The initial search identified 182 references corresponding to 117 original studies (65 were duplicated) and 17 references corresponding to reviews and reports from Agencies for Health Technology Assessments. Eight publications fulfilled the inclusion criteria. One of them was excluded because it studied a different procedure, denervation by radiofrequency of the ramus communicans,¹⁶ and 2 of them corresponded to follow-up results at 12 months¹⁷ and 24 months⁷ of the same study, so that only the study including the longest follow-up was included. Therefore, 6 studies, including a total of 283 patients,

were included in this systematic review. Characteristics of the included studies are shown in Table 1.

Two RCTs assessed the efficacy of IDET *versus* placebo,^{8,9} and 1 nonrandomized trial assessed its effectiveness *versus* a rehabilitation program consisting of physical therapy, exercise, education, and counseling.⁷ One RCT assessed the efficacy of PIRFT *versus* placebo,¹² and another one compared PIRFT techniques, which were different in terms of temperature (70°C *vs.* 80°C) and duration (120–360 seconds *vs.* 90 seconds).¹⁰ The last nonrandomized trial was a prospective matched control trial comparing IDET *versus* PIRFT.¹¹

In all the studies, the population was composed of patients with chronic LBP who had not improved with conservative treatment, although for a variable duration (from only 6 weeks,⁸ to 2 years¹⁰). According to the inclusion criteria, all included studies tried to establish the discogenic origin of pain. To that purpose, 5 studies used provocative lumbar discography and 1 used analgesic discography.¹²

The mean age of included patients was around 40 years and, in all studies but one,⁹ more women than men were included. Baseline pain intensity was approximately 7 points on a 10 points visual analog scale,¹⁸ although this was not reported in 1 study.⁹ Baseline LBP-related disability was over 40% in Oswestry Disability Scale,¹⁹ although in 1 study⁸ mean Oswestry Disability Scale score was 10 points lower than in the rest of the studies. Most studies had a short follow-up period (from 8 weeks to 6 months) as shown in Table 1. In 2 studies, there were a high number of losses to follow-up, which were excluded from the analysis (8 of 64 in Pauza, 2004; and 8 of 36 in Bogduk, 2002).^{7,8} There were no losses in 1 study.¹² They were not reported in another one,¹¹ and there were a small number in the rest of them (2 of 57 in Freeman, 2005; and 2 of 39 in Erçelen, 2003).^{9,10}

Methodologic Quality

The assessment of methodologic quality is shown in Table 2. Among randomized clinical trials, randomization was assigned by computer in 3 studies.^{8,10,12} Randomization was centralized in 1 RCT,⁸ and possibly in another one.¹² There is no detailed information about randomization in Freeman (2005), where only “sealed envelopes” are mentioned.⁹ As to assessment, 3 studies were double-blind.^{8,9,12} Masking is not specified in another study,¹⁰ and the 2 nonrandomized trials were open.^{7,11} Therefore, only 3 studies had a reasonable methodologic design, with adequate methods of randomization and allocation concealment.^{8,9,12}

Outcome Results

Results of the 6 included studies are summarized in Tables 3 to 8. All the studies included pain and disability as outcomes. One study⁹ used the Low Back Outcome Score²⁰ to assess pain, and another one¹¹ used the Pain Disability Index to measure disability.²¹ All the other studies used visual analog scale¹⁸ for pain and the Oswestry Disability Scale¹⁹ for disability.

Table 1. Characteristics of Included Studies

Study ID	Methods	Participants	Interventions	Outcomes
Barendse (2001)	RCT randomization assigned by computer; patients and investigators unaware of group allocation	Twenty-eight patients with chronic nonspecific low back pain >1 yr; 1 level of discogenic pain (L4–L5 or L5–S1); age range, 30–65 yr; no significant differences at baseline; diagnosed by analgesic discography	Treatment group: PIRFT 90 s at 70°C; Control group: sham intervention (identical process with no RF applied); no data about cointerventions	Follow-up: 8 wk; rating of pain by VAS; global perceived effect by the patient analgesic intake; disability assessed by the Oswestry Disability Scale; quality of life (COOP-WONCA chart); therapeutic success: 2-point reduction on the VAS scale and at least 50% reduction on global perceived effect
Bogduk-Karasek (2002)	Prospective cohort study (convenience sample study); all patients requested IDET; the control group consisted of patients in which the insurance denied the procedure and were actively pursuing litigation to have it done	Fifty-three patients with low back pain for 3 mo and no response to conservative treatment; no evidence of disc prolapse, neurologic disease, tumor, or infection; discography performed at 3 levels and confirmed by CT	Treatment group: IDET with electrode heated at 90°C, 80°C, or 85°C according to patient tolerance; Control group: rehabilitation program (involving physical therapy, strengthening and conditioning exercises, education and counseling); cointervention: cefazolin given as antibiotic prophylaxis in the treatment group; also, postoperative rehabilitation program	Follow-up: 3 mo (Karasek, 2000) and 12–24 mo (Bogduk, 2002); rating of pain by VAS; % patients with pain improving; return to work; use of opioid analgesics; therapeutic success: 50% reduction of pain (or no pain) + able to work + no opioids
Ercelen (2003)	Randomization assigned by RCT computer; no details about allocation concealment	Thirty-nine patients with chronic low back pain with no response to conservative treatment for 2 yr; normal neurologic examination; no nerve compression on MRI; provocative discography at 2 levels	PIRFT (heating 80°C during 120 s vs. 360 s); cointervention: 1 g cefazolin given IV in both groups	Follow-up: 6 mo; rating of pain by VAS; disability assessed by the Oswestry Disability Scale
Freeman (2005)	RCT randomization by sealed envelope; both subject and surgeon blinded	Fifty-seven patients with chronic discogenic low back pain, functional disability, degenerative disc disease on MRI, failed conservative management; provocative lumbar discography and CT	IDET group: heating commenced at 65°C rising over 12.5 min to 90°C and held for 4 min; Control group: the catheter was not connected to a generator; cointervention: 1 g IV cefazolin + 100 mg cefazolin via the introducer needle in both groups	Follow-up: 6 mo; back pain: low back outcome score; disability: Oswestry Disability Scale; depression: Zung Depression Index; Somatic: MSPQ; QoL: SF-36; sitting tolerance; work tolerance; medication; any neurologic deficit
Kapural (2005)	Prospective matched control trial; patients matched for age, sex, weight, smoking history, manual labor, and no. of intravertebral discs treated; physician who performed all the annuloplasties not blinded	Forty-two patients with low back pain unresponsive to conservative treatment for >6 mo; no evidence of compressive radiculopathy, no prior surgery, no signs of lumbar stenosis, no psychological issues; evidence of single-level or two-level disc disease at MRI; provocation discography	IDET group: rising temperature inside the disc from 65°C to 90°C and maintained for 4 min; PIRFT group: 55°C for 4 min, 60°C for 5 min, and 65°C for 5 min; no data about cointerventions	Follow-up: 12 mo; rating of pain by VAS; disability assessed by the Pain Disability Index
Pauza (2004)	RCT randomization assigned by computer; patients and investigators unaware of group allocation	Sixty-four patients with low back pain, unresponsive to conservative treatment for 6 wk; low back pain worse sitting or standing, relieved by lying down; Beck depression scale <20; no surgical interventions previous 3 mo; <20% disc height narrowing; provocation discography and CT	IDET group: catheter heated to 90°C over 17 min; Control group: sham intervention with needle introduced, monitor demonstrated the passage of an electrode and generator noises; cointervention: IV prophylactic antibiotic + cephazolin 0.056 mg/mL or gatifloxacin 0.33 mg/mL injected into the disc; rehabilitation program both groups	Follow-up: 6 mo; rating of pain by VAS; QoL: SF-36; disability assessed by the Oswestry Disability Scale; reduction of pain: % patients with 25%, 50%, and 75% relief of pain

Table 2. Methodologic Quality of Included Studies

Item	Barendse (2001)	Bogduk-Karasek (2002)	Ercelen (2003)	Freeman (2005)	Kapural (2005)	Pauza (2004)
Was randomization performed?	Yes	No	Yes	Yes	No	Yes
Was treatment allocation concealed?	Yes	No	Not specified	Yes	No	Yes
Were the groups similar at baseline regarding the most important prognostic indicators?	Yes	No†	Yes	Yes	No#	Yes**
Was the care provider blinded?	Yes	No	Not specified	Yes	No	Yes
Were the cointerventions avoided or comparable?	No data	No‡	Comparable	Comparable	No data	Comparable
Was the compliance rate acceptable in all groups?	Yes	No§	Yes	Yes	Yes	Yes
Was the patient blinded?	Yes	No	Not specified	Yes	No	Yes
Was the outcome assessor blinded?	Yes	No	Not specified	Yes	No	Yes
Was the withdrawal/dropout rate unlikely to cause bias?	Yes	Yes	Yes	Yes	Yes	Not††
Was the timing of the outcome assessment in both groups comparable?	No*	No¶	Yes	Yes	Yes	Yes

*Assessment at 8 wk both groups. Success patients also assessed at 3, 6, and 12 months.

†Not enough information provided.

‡Antibiotic prophylaxis with cefazolin and graded reactivation program only in IDET group.

§No evidence that the control group took the rehab conscientiously as they were all still waiting for the IDET through the study period and according to the authors were pursuing litigation against the insurance carriers to have the IDET.

||No withdrawals/dropouts at 3 months.

¶Study designed to perform a direct comparison at 3 months and prospective follow-up at 12 and 24 months of the IDET group only. Nevertheless, the control group was also assessed at 12 and 24 months, when possible.

#No statistically significant differences.

**There were some differences re previous physiotherapy treatment (70% sham group vs. 41% treatment group) and role-physical of the SF-36 (mean 39 in IDET group vs. 16 in sham group).

††There were 8 losses out of 64 (5 in the IDET group and 3 in the sham group) that were not included in the analysis.

In addition, 1 randomized trial¹² and a nonrandomized trial⁷ measured pain reduction using a dichotomous outcome consisting of the percentage of patients who reached a certain degree of pain relief.

Other outcomes that were used in different studies were therapeutic success (using different criteria),^{7,12} quality of life (measured with SF-36)²²⁻²⁴ in 1 study,⁸ and with the COOP-WOMCA charts^{25,26} in another one.¹² Global effect perceived by the patient and return to work were also assessed in 1 RCT.¹²

Table 3. Results of Pauza (2004): Follow-up at 6 Months

Outcome Measure	IDET (n = 32)		Sham (n = 24)		P
	Mean	SD	Mean	SD	
VAS for pain (0-10)					
Pretreatment	6.6	1.4	6.5	1.9	0.758
6 mo	4.2	2.6	5.4	2.7	0.089
Change	2.4	2.3	1.1	2.6	0.045
SF-36: bodily pain (0-100)					
Pretreatment	36	12	35	12	0.765
6 mo	53	19	44	20	0.085
Change	17	19	9	15	0.086
SF-36: physical functioning (0-100)					
Pretreatment	56	24	49	21	0.236
6 mo	71	22	60	24	0.079
Change	15	27	11	17	0.548
Oswestry disability scale (0-100)					
Pretreatment	31	10	33	11	0.485
6 mo	20	12	28	15	0.023
Change	11	11	4	12	0.050

In both RCTs on IDET *versus* placebo,^{8,9} pain, disability, and quality of life were assessed for 6 months. In the first one,⁸ there were small differences in favor of IDET for pain and disability (although the difference on the latter was clinically irrelevant),^{27,28} while there was none in the highest quality recent one.⁹ The latter also assessed depression, sitting and work tolerance, medication and neurologic deficit, finding no differences between IDET and placebo.

In the RCT on PIRFT *versus* placebo,¹² pain, disability, quality of life, global effect perceived by the patient, therapeutic success, and analgesic intake were assessed for 8 weeks, showing no differences between groups.

In the RCT comparing different methods of PIRFT,¹⁰ pain and disability were assessed for 6 months, and no differences were found between groups. As compared with baseline values, pain and disability improved at month 1, but this effect disappeared from month 2 onwards. In the nonrandomized trial comparing IDET *versus* PIRFT,¹¹ a better evolution of pain and disability at 1 year was observed in the IDET group.

In the nonrandomized trial comparing IDET and a rehabilitation program,⁷ the proportion of patients with a ≥50% pain reduction was higher in the IDET group, both 12 and 24 months after the procedure.

Side Effects

No side effects were detected in 2 RCTs on IDET,^{8,12} although in a third one,⁹ a transient radiculopathy (<6 weeks), was reported in 4 patients in the IDET group and 1 in the sham group. In the RCT comparing 2 methods of

Table 4. Results of Freeman (2005): Follow-up at 6 Months

Characteristic	Baseline				6 Months			
	IDET (n = 38)		Placebo (n = 19)		IDET (n = 36)		Placebo (n = 19)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Low back outcome score*	39.51	5.25	36.71	3.00	38.31	3.61	37.45	1.60
Oswestry Disability Index†	41.42	14.80	40.74	11.84	39.77	16.28	41.58	11.29
Zung Depression Index‡	41.03	6.13	40.42	10.04	41.39	4.46	40.82	7.72
MSPQ§	8.22	5.16	6.84	5.88	8.67	6.09	8.67	4.37
SF-36								
Physical functioning	41.86	23.01	35.00	15.37	44.72	24.20	36.58	20.14
Role-physical	13.82	30.03	5.26	10.47	20.83	34.59	13.89	23.04
Pain index	33.13	15.97	24.42	13.45	38.28	21.37	31.47	15.29
General health perceptions	65.29	19.21	60.33	17.79	61.44	22.68	64.16	19.29
Vitality	38.86	21.69	45.53	16.32	37.08	25.22	45.79	21.16
Social functioning	41.12	27.86	44.08	19.71	45.14	30.80	43.42	20.14
Role-emotional	46.49	42.83	46.30	42.99	42.59	44.09	38.89	36.60
Mental health index	55.89	21.30	64.00	13.68	52.22	23.11	61.26	19.42
Standard physical component scale	32.58	8.07	26.90	5.74	35.10	8.70	30.40	6.15
Standard mental component scale	40.34	12.76	44.77	8.29	38.16	13.29	43.05	11.07

*The higher the score, the better the outcome.
 †The higher the score, the worse the outcome.
 ‡The higher the score, the greater the degree of depression.
 §The higher the score, the worse the outcome.
 ||The higher the score, the better the outcome.

PIRFT,¹⁰ 1 patient had discitis. Side effects were not assessed in the nonrandomized trials.^{7,11}

Discussion

Results from this systematic review do not support the efficacy or effectiveness of percutaneous thermocoagulation intradiscal techniques.²⁹

In the case of PIRFT, the evidence suggests that it is an ineffective technique. The only available high-quality RCT shows that the procedure is not better than placebo,¹² and the other available RCT¹⁰ shows that PIRFT leads to re-

sults that are similarly poor with no difference if the heating temperature is applied for 2 or for 6 minutes.

In the case of IDET, several previous case reports suggested that the procedure might be effective.^{14,30-39} However, those reports derive from data registers so they could not take into account the effect of regression to the mean,⁴⁰ the natural history of the condition, the placebo effect and other potential confounders, such as cointerventions and other mechanical and psychosocial factors having shown to influence patients' evolution. In addition, most of them come from promoters of the proce-

Table 5. Results of Bogduk-Karasek (2002): Follow-up at 3, 6, 12, and 24 Months

	Patients to Which IDET Was Denied (Control Group)		Patients to Which IDET Was Applied (Intervention Group)		P
	N	Median VAS for Pain	N	Median VAS for Pain	
Before treatment	17	8.0 (5-8)	36	8.0 (7-9)	0.071
At 3 mo	17	8.0 (7-8)	36	3.5 (1-5)	0.000
At 6 mo	3.0 (1-6)				
At 12 mo	12	7.5 (5-8)	35	3.0 (1-7)	0.005
At 24 mo	10	7.5 (4-8)	35	3.0 (1-7)	0.028
	<u>N</u>	<u>IDET Group</u>	<u>N</u>	<u>Control Group</u>	
Therapeutic success					
Criteria 1: 50% reduction VAS + return to work + no use of opioids					
12 mo	35	21 (60%)	10	1 (10%)	
24 mo	35	18 (54%)	10	1 (10%)	
P < 0.013					
Criteria 2: Complete relief of pain + return to work or remaining at work + no use of opioids					
12 mo	35	8 (23%)	10	0	
24 mo	35	7 (20%)	10	0	

Differences not statistically significant

Table 6. Results of Barendse (2001): Follow-up at 8 Weeks

	Mean Sham Group	Mean PIRFT Group	Difference Unadjusted (90% CI)	Difference Adjusted (90% CI)
Change in VAS: mean	-1.14	-0.61	-0.53 (-1.95 to 0.89)	1.25 (-0.55 to 3.06)
Change in VAS: high	-1.21	-1.35	0.15 (-1.32 to 1.61)	0.67 (-1.11 to 2.46)
Change in VAS: low	-0.07	0.27	-0.34 (-1.79 to 1.11)	0.55 (-1.42 to 2.52)
Global perceived effect	0.21	0.09	0.12 (-0.78 to 1.03)	-0.18 (-1.28 to 0.91)
Change in impairment according to Waddell	0.29	0.00	0.52 (-0.65 to 1.22)	-0.51 (-1.73 to 0.70)
Change in no. of analgesic tablets per 4 days	0.43	-1.38	1.81 (-2.15 to 5.77)	-2.26 (-7.04 to 2.53)
Change in Oswestry Disability Scale	-4.93	-2.62	-2.31 (-10.08 to 5.45)	3.28 (-7.54 to 14.11)
Change in Coop/Wonca	-0.21	-1.85	1.63 (-0.49 to 3.75)	-1.06 (-3.88 to 1.77)

cedure, so that potential conflicts of interest cannot be ruled out.

A first low-quality trial with 53 patients suggested that IDET improved a higher proportion of patients than a rehabilitation program.⁷ However, that was a nonrandomized trial, the size of the unspecific (placebo) effect triggered by the 2 interventions was likely to be different, compliance in the rehabilitation group was not measured, and assessment was not blinded. In addition, baseline patients' characteristics were different in terms of age and levels of painful disc, and other potential confounders were not analyzed either. Finally, all patients included in the study requested IDET, and the "control" group was composed of those in which the procedure was denied by their insurance and the patients were actively pursuing litigation to have the IDET done. Not surprisingly, results from this study were inconsistent with those from proper RCTs.

Indeed, 2 RCTs were later performed, leading to contradictory results. According to the highest-quality RCT, IDET is not better than placebo for any variable.⁹ On the contrary, the other available RCT⁸ suggests that IDET is slightly better than placebo for pain and disability, although the size of the effect on the latter is clinically irrelevant.^{27,28} However, positive results in this study may be explained by reasons other than potential effectiveness of the procedure. IDET and control groups were not comparable at baseline, and all those differences as well as the cointerventions favored the IDET group. In addition, those differences were not taken into account at the analysis phase, and multiple *post hoc* comparisons were made. Moreover, there were losses to follow-up (5

of 37 patients in the IDET group, and 3 of 37 in the placebo group). Some of those losses corresponded to patients with "reinjuries," and no analysis by intention to treat was performed, but only "per protocol." Finally, the main difference in favor of IDET was the higher proportion of patients in that group achieving a "successful" reduction in pain; however, the decision on what magnitude of pain reduction was to be considered a "success" was made *post hoc*. Interestingly enough, the high-quality RCT in which the criterion of "success" was defined before the study found no differences between IDET and placebo.⁹ Taken together, the available evidence is at best conflictive and does not allow to draw firm conclusions on the efficacy of IDET.

A small low-quality trial suggested that IDET improves pain and disability more than PIRFT does.¹¹ This was a nonrandomized trial with 42 patients that were matched for age, sex, weight, smoking history, manual labor, and number of intervertebral discs treated. Taking into account limitations derived from its design, and that there is no consistent evidence that both IDET and PIRFT are better than placebo, it is impossible to draw firm conclusions on the efficacy or effectiveness of IDET based on this study.

Side effects from intradiscal thermocoagulation techniques were not searched for in the 2 nonrandomized trials,^{7,11} and were not reported in 2 randomized trials.^{8,12} Among the other studies included in this review, 1 reported a transient radiculopathy in 4 participants of the IDET group and 1 of the control group,⁹ and in another one a patient was excluded because of discitis.¹⁰ In addition, more severe side effects were described in previous case reports, such as catheter breakage, verte-

Table 7. Results of Ercegen (2003): Follow-up at 6 Months

	Group PIRFT 120			Group PIRFT 360		
	VAS (mean ± SD)	Pain Relief (%)	ODS (%)	VAS (mean ± SD)	Pain Relief (%)	ODS (%)
Preprocedure	6.73 ± 1.55		42.4 ± 9.3	6.27 ± 1.31		41.9 ± 10.2
Immediate	1.21 ± 0.797*	82.02		0.94 ± 1.05*	85.00	
1 wk	2.68 ± 0.94*	60.17		2.55 ± 1.24*	59.33	
2 wk	3.15 ± 0.76*	53.19		3.05 ± 1.21*	49.64	
1 mo	3.36 ± 0.89*	46.66	26.3 ± 10.9*	3.33 ± 0.97*	46.88	24.1 ± 12.1*
2 mo	5.26 ± 2.40	21.84		4.94 ± 2.36	21.21	
3 mo	5.31 ± 2.35	21.09		5.00 ± 2.61	20.25	
6 mo	5.42 ± 2.43	19.46	38.8 ± 14.4	4.83 ± 2.14	22.96	37.6 ± 13.9

*P < 0.05.

Table 8. Results of Kapural (2005): Follow-up at 12 Months

Time	Estimated Mean Difference in PDI (IDET-PIRFT)	95% CI		P
		Lower Level	Upper Level	
Preprocedure	7.2	-3.0	17.4	0.16
2 wk	-0.95	-11.2	9.2	0.85
2 mo	-5.2	-15.4	5.0	0.31
3 mo	-17.3	-27.5	-7.1	<0.001
6 mo	-16.8	-27.0	-6.6	0.001
9 mo	-24.2	-34.4	-14.0	<0.001
1 yr	-21.8	-32.0	-11.6	<0.001

Values are differences in mean pain disability index (PDI) between the 2 groups.

bral osteonecrosis, and cauda equina syndrome.⁴¹⁻⁴⁴ This suggests that intradiscal thermocoagulation techniques have potentially severe side effects, although its exact rate is currently unknown.

The evidence does not support the effectiveness of IDET or PIRFT, and those procedures are associated with potentially serious side effects.⁴¹⁻⁴⁴ However, more than 75,000 IDET procedures have been performed in the United States alone since 1998.⁴⁵ For ethical, clinical, and legal reasons, this behavior seems inappropriate. It has been pointed out that assessment of efficacy, safety, and effectiveness of nonpharmacological technologies are usually less rigorous than the one of drugs.⁴⁶ In addition, economic interests seem to influence at least the pace and the characteristics of the assessment, approval and implementation processes, and industry support is associated with differences in reported results.^{46,47} As in the case of percutaneous intradiscal thermocoagulation techniques, the use of other invasive procedures for treating LBP has spread in the absence of high-quality evidence on its efficacy or effectiveness. This may lead to inadequate practice, inefficient cost expenditures, and patients' exposure to unnecessary risks.^{48,49} Chronic patients with LBP represent a difficult population, so it is possible that in some cases clinicians' desperate search for effective techniques may account for their willingness to try any new procedure for that condition as soon as it is commercially available. However, profit seems to be the driving force in the development and spread of spinal invasive treatments,⁴⁸⁻⁵⁰ and the premature use of potentially risky procedures with no sound evidence supporting their effectiveness, can hardly be seen as a positive way to help patients. Therefore, clinicians treating patients with LBP should refrain from using new technologies based on promising results deriving from case reports or nonrandomized trials. Instead, they should wait until high-quality RCTs confirm the efficacy and effectiveness of technologies before using them in the clinical environment.

It is uncertain that provocation and analgesic discography are reliable methods to diagnose discogenic LBP, as they are prone to produce false-positive responses.⁵¹⁻⁵³

Some studies have suggested that the effectiveness of discography may improve by using also MRI to detect high-intensity zones in the affected disc,⁵⁴⁻⁵⁷ but high-intensity zones and pain on injection also correlate in asymptomatic subjects. Nowadays, no procedure can reliably identify patients in which pain originates from the disc, which makes it impossible to perform trials on treatment methods focusing on that hypothetical condition.

The available evidence does not currently support the need to perform further studies on PIRFT. In the case of IDET, RCTs are consistent in showing no relevant effect on disability and quality of life. The highest-quality RCT shows no effect on pain, and the positive effect on that variable shown in a methodologically flawed study can be explained by reasons other than its potential effectiveness. This seems a weak basis for further studies on IDET. However, if such studies were to be envisaged, they would only make sense after having validated procedures to identify accurately patients with discogenic LBP. Those potential RCTs should be performed in different settings, with very large populations accurately identified as having discogenic LBP, and should assess efficacy *versus* a sham procedure before comparing its effectiveness *versus* other procedures having consistently shown their effectiveness, such as cognitive behavioral therapy,⁵⁸ or neuroreflexotherapy.⁵⁹ In those potential trials, data regarding safety should be systematically collected, since some severe adverse events have been noticed in several case reports and were not detected in the small RCTs performed so far. In the absence of consistent positive results deriving from such trials, the clinical use of percutaneous thermocoagulation intradiscal techniques can currently not be recommended.

■ Key Points

- The intervertebral disc has been proposed as a source of pain in a major proportion of patients with low back pain. Two percutaneous thermocoagulation intradiscal techniques have been described to treat discogenic low back pain: IDET and PIRFT.
- Although previous case reports and nonrandomized trials suggested positive results, results from randomized clinical trials show that PIRFT is not effective for the treatment of discogenic low back pain. Regarding IDET, 1 RCT shows a positive effect only on pain severity, while the best quality RCT shows no effect on any variable.
- The available evidence does not support the clinical use of PIRFT or IDET, and potentially serious adverse effects have been reported.
- These results confirm that case reports and nonrandomized clinical trials are not useful to assess the effectiveness of treatments. Clinical use of procedures should be postponed until results from high-quality randomized clinical trials are available.

Acknowledgments

The authors thank Dr. Sera Tort for editing of this manuscript.

References

- Singh K, Ledet E, Carl A. Intradiscal therapy: a review of current treatment modalities. *Spine* 2005;30(suppl):20–6.
- Hildebrandt J. Relevance of nerve blocks in treating and diagnosing low back pain: is the quality decisive? *Schmerz* 2001;15:474–83.
- O'Neill CW, Kurgansky ME, Derby R, et al. Disc stimulation and patterns of referred pain. *Spine* 2002;27:2776–81.
- Bogduk N. The causes of low back pain. *Med J Aust* 1992;156:151–3.
- Bogduk N. The lumbar disc and low back pain. *Neurosurg Clin North Am* 1991;2:791–806.
- Donelson R, Aprill C, Medcalf R, et al. A prospective study of centralization of lumbar and referred pain: a predictor of symptomatic discs and anular competence. *Spine* 1997;22:1115–22.
- Bogduk N, Karasek M. Two-year follow-up of a controlled trial of intradiscal electrothermal annuloplasty for chronic low back pain resulting from internal disc disruption. *Spine J* 2002;2:343–50.
- Pauza KJ, Howell S, Dreyfuss P, et al. A randomized, placebo-controlled trial of intradiscal electrothermal therapy for the treatment of discogenic low back pain. *Spine J* 2004;4:27–35.
- Freeman BJC, Fraser RD, Cain CMJ, et al. A randomized, double-blind, controlled trial: intradiscal electrothermal therapy versus placebo for the treatment of chronic discogenic low back pain. *Spine* 2005;30:2369–77.
- Erçelen O, Bulutçu E, Oktenoglu T, et al. Radiofrequency lesioning using two different time modalities for the treatment of lumbar discogenic pain: a randomized trial. *Spine* 2003;28:1922–7.
- Kapural L, Hayek S, Malak O, et al. Intradiscal thermal annuloplasty versus intradiscal radiofrequency ablation for the treatment of discogenic pain: a prospective matched control trial. *Pain Med* 2005;6:425–31.
- Barendse GA, van Den Berg SG, Kessels AH, et al. Randomized controlled trial of percutaneous intradiscal radiofrequency thermocoagulation for chronic discogenic back pain: lack of effect from a 90-second 70°C lesion. *Spine* 2001;26:287–92.
- Carragee EJ, Tanner CM, Yang B, et al. False-positive findings on lumbar discography: reliability of subjective concordance assessment during provocative disc injection. *Spine* 1999;24:2542–7.
- Saal JS, Saal JA. Management of chronic discogenic low back pain with a thermal intradiscal catheter: a preliminary report. *Spine* 2000;25:382–8.
- Van Tulder MW, Assendelft WJJ, Koes BW, et al. Method guidelines for systematic reviews in the Cochrane Collaboration back review group for spinal disorders. *Spine* 1997;22:2323–30.
- Oh WS, Shim JC. A randomized controlled trial of radiofrequency denervation of the ramus communicans nerve for chronic discogenic low back pain. *Clin J Pain* 2004;20:55–60.
- Karasek M, Bogduk N. Twelve-month follow-up of a controlled trial of intradiscal thermal annuloplasty for back pain due to internal disc disruption. *Spine* 2000;25:2601–7.
- Huskinson EC. Measurement of pain. *Lancet* 1974;2:1127–31.
- Fairbank JC, Couper J, Davies JB, et al. The Oswestry Low Back Pain Disability Questionnaire. *Physiotherapy* 1980;66:271–3.
- Greenough CG, Fraser RD. Assessment of outcome in patients with low back pain. *Spine* 1992;17:36–41.
- Pollard CA. Preliminary validity study of the pain disability index. *Percept Mot Skills* 1984;59:974.
- Ware JE, Sherbourne CD. The MOS 36-item Short-Form Health Survey (SF-36): I. Conceptual framework and item selection. *Med Care* 1992;30:473–83.
- McHornery CA, Ware JE, Rogers WH, et al. The validity and relative precision of MOS Short- and Long-Form Health Status Scales and Dartmouth COOP charts: results from Medical Outcomes Study. *Med Care* 1992;30(suppl):MS253–65.
- McHornery CA, Ware JE, Raczek AE. The MOS 36-item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care* 1993;31:247–63.
- Nelson EC, Wasson J, Kirk J. Assessment of function in routine clinical practice: description of the COOP chart method and preliminary findings. *J Chron Dis* 1987;40(suppl):55–64.
- Scholten JHG, van Weel C. *Functional Status Assessment in Family Practice*. Meditekst: Lelystad, 1992.
- Bombardier C, Hayden J, Beaton DE. Minimal clinically important difference in low back pain outcome measures. *J Rheumatol* 2001;28:431–8.
- Häag O, Fitzell P, Nordwall A. The clinical importance of changes in outcome scores after treatment for chronic low back pain. *Eur Spine J* 2003;12:12–20.
- Higgins JPT, Green S, eds. *Cochrane Handbook for Systematic Reviews of Interventions*, 4.2.5 [updated May 2005]. In: *The Cochrane Library*, Issue 3, 2005. Chichester, UK: John Wiley & Sons, 2005.
- Saal JA, Saal JS. Intradiscal electrothermal treatment for chronic discogenic low back pain: as prospective outcome study with minimum 1-year follow-up. *Spine* 2000;25:2622–7.
- Singh V. Intradiscal electrothermal therapy: a preliminary report. *Pain Physician* 2000;3:367–73.
- Derby R, Eek B, Chen Y, et al. Intradiscal electrothermal annuloplasty (IDET): a novel approach for treating chronic discogenic back pain. *Neuro-modulation* 2000;3:5–9.
- Saal SA, Saal JS. Intradiscal electrothermal therapy for chronic discogenic low back pain: prospective outcome study with a minimum 2-year follow-up. *Spine* 2002;27:966–74.
- Endres SM, Fiedler GA, Larson KL. Effectiveness of intradiscal electrothermal therapy in increasing function and reducing chronic low back pain in selected patients. *Wis Med J* 2002;101:31–4.
- Gerszten PC, Welch WC, McGrath PM, et al. A prospective outcomes study of patients undergoing intradiscal electrotherapy (IDET) for chronic low back pain. *Pain Physician* 2002;5:360–4.
- Spruit M, Jacobs WC. Pain and function after intradiscal electrothermal treatment (IDET) for symptomatic lumbar disc degeneration. *Eur Spine J* 2002;11:589–93.
- Lutz C, Lutz GE, Cooke PM. Treatment of chronic lumbar discogenic pain with intradiscal electrothermal therapy: a prospective outcome study. *Arch Phys Med Rehabil* 2003;84:23–8.
- Mekhail N, Kapural L. Intradiscal thermal annuloplasty for discogenic pain: an outcome study. *Pain Practice* 2004;4:84–90.
- Finch PM, Price LM, Drummond PD. Radiofrequency heating of painful annular disruptions: one-year outcomes. *J Spinal Disord Tech* 2005;18:6–13.
- Morton V, Torgerson DJ. Effect of regression to the mean on decision making in health care. *BMJ* 2003;326:1083–4.
- Hsia AW, Isaac K, Katz JS. Cauda equina syndrome from intradiscal electrothermal therapy. *Neurology* 2000;55:320.
- Djurasovic M, Glassman SD, Dimar JR, et al. Vertebral osteonecrosis associated with the use of intradiscal electrothermal therapy: a case report. *Spine* 2002;27:E 325–8.
- Wetzel FT. Cauda equina syndrome from intradiscal electrothermal therapy. *Neurology* 2001;56:1607.
- Konno S, Olmarker K, Byrod G, et al. Acute thermal nerve root injury: the European Spine Society AcroMed Prize 1994. *Eur Spine J* 1994;3:299–302.
- http://www.idetprocedure.com/2000_professional/2100_about.html Accessed July 6, 2006.
- Deyo RA. Gaps, tensions, and conflicts in the FDA approval process: implications for clinical practice. *J Am Board Fam Pract* 2004;17:142–9.
- Shah RV, Todd JA, Bruegel-Sanchez V, et al. Industry support and correlation to study outcome for papers published in Spine. *Spine* 2005;30:1099–104.
- Deyo RA, Racy DT, Kreuter W, et al. United States trends in lumbar fusion surgery for degenerative conditions. *Spine* 2005;30:1441–5.
- Deyo R, Patrick D. *Hope or Hype*, 1st ed. New York: AMACOM, 2005.
- Weiner AK, Levi BH. The profit motive and spine surgery. *Spine* 2004;29:2588–91.
- Carragee EJ, Paragioudakis SJ, Khurana S. 2000 Volvo Award winner in clinical studies: lumbar high-intensity zone and discography in subjects without low back problems. *Spine* 2000;25:2987–92.
- Smith BM, Hurwitz EL, Solsberg D, et al. Interobserver reliability of detecting lumbar intervertebral disc high-intensity zone on magnetic resonance imaging and association of high-intensity zone with pain and anular disruption. *Spine* 1998;23:2074–80.
- Schwarzer AC, Aprill CN, Derby R, et al. The prevalence and clinical features of internal disc disruption in patients with chronic low back pain. *Spine* 1995;20:1878–83.
- Milette PC, Fontaine S, Lepanto L, et al. Differentiating lumbar disc protrusions, disc bulges, and discs with normal contour but abnormal signal intensity: magnetic resonance imaging with discographic correlations. *Spine* 1999;24:44–5.
- Ricketson R, Simmons JW, Hauser BO. The prolapsed intervertebral disc: the high-intensity zone with discography correlation. *Spine* 1996;21:2758–62.

56. Schellhas KP, Pollei SR, Gundry CR, et al. Lumbar disc high-intensity zone: correlation of magnetic resonance imaging and discography. *Spine* 1996;21:79–86.
57. Vanharanta H, Ohnmeiss DD, Aprill CN. Vibration pain provocation can improve the specificity of MRI in the diagnosis of symptomatic lumbar disc rupture. *Clin J Pain* 1998;14:239–47.
58. Ostelo RWJG, van Tulder MW, Vlaeyen JWS, et al. Behavioural treatment for chronic low-back pain. In: *The Cochrane Database of Systematic Reviews*, Issue 1. 2005.
59. Urrútia G, Burton AK, Morral A, et al. Neuroreflexotherapy for non-specific low-back pain. *The Cochrane Database of Systematic Reviews*, Issue 2. 2004.

■ Appendix

Appendix. Search strategy

The search strategy was based on the following terms in free text:

- #1 Intradiscal electrothermal annuloplasty[tw]
- #2 intradiscal electrothermal therapy[tw] OR IDET[tw]
- #3 percutaneous intradiscal radiofrequency thermo-coagulation[tw] OR PIRFT[tw]
- #4 or 1–3

LITERATURE REVIEW

Surgery Versus Conservative Treatment for Symptomatic Lumbar Spinal Stenosis

A Systematic Review of Randomized Controlled Trials

Francisco M. Kovacs, MD, PhD,*|| Gerard Urrútia, MD,†‡§|| and José Domingo Alarcón, MD¶||

Study Design. Systematic review.**Objective.** To compare the effectiveness of surgery versus conservative treatment on pain, disability, and loss of quality of life caused by symptomatic lumbar spinal stenosis (LSS).**Summary of Background Data.** LSS is the most common reason for spine surgery in persons older than 65 years in the United States.**Methods.** Randomized controlled trials (RCTs) comparing any form of conservative and surgical treatment were searched in CENTRAL, MEDLINE, EMBASE, and TripDatabase databases until July 2009, with no language restrictions. Additional data were requested from the authors of the original studies. The methodological quality of each study was assessed independently by two reviewers, following the criteria recommended by the Cochrane Back Review Group. Only data from randomized cohorts were extracted.**Results.** A total of 739 citations were reviewed. Eleven publications corresponding to five RCTs were included. All five scored as high quality despite concerns deriving from heterogeneity of treatment, lack of blinding, and potential differences in the size of the placebo effect across groups. They included a total of 918 patients in whom

conservative treatments had failed for 3 to 6 months, and included orthosis, rehabilitation, physical therapy, exercise, heat and cold, transcutaneous electrical nerve stimulation, ultrasounds, analgesics, nonsteroidal anti-inflammatory drugs, and epidural steroids. Surgical treatments included the implantation of a specific type of interspinous device and decompressive surgery (with and without fusion, instrumented or not). In all the studies, surgery showed better results for pain, disability, and quality of life, although not for walking ability. Results of surgery were similar among patients with and without spondylolisthesis, and slightly better among those with neurogenic claudication than among those without it. The advantage of surgery was noticeable at 3 to 6 months and remained for up to 2 to 4 years, although at the end of that period differences tended to be smaller.

Conclusion. In patients with symptomatic LSS, the implantation of a specific type of device or decompressive surgery, with or without fusion, is more effective than continued conservative treatment when the latter has failed for 3 to 6 months.**Key words:** lumbar spinal stenosis, systematic review, randomized controlled trial, surgery, conservative treatment. **Spine 2011; 36:E1335–E1351**

From the *Departamento Científico, Fundación Kovacs, Palma de Mallorca, Spain; †Iberoamerican Cochrane Center-Servei d'Epidemiologia Clínica i Salut Pública, Institut d'Investigació Biomèdica Sant Pau, Barcelona, Spain; ‡CIBER Epidemiología y Salud Pública, Barcelona, Spain; §Public Health and Research Methodology Programme, Universitat Autònoma de Barcelona, Barcelona, Spain; ¶Iberoamerican Cochrane Network, Universidad Surcolombiana, Neiva, Colombia; and ||Red Española de Investigadores en Dolencias de la Espalda, Paseo Mallorca 36, 07012 Palma de Mallorca, Spain.

Acknowledgement date: June 1, 2010. First Revision date: August 16, 2010. Second Revision date: October 12, 2010. Third Revision date: November 17, 2010. Acceptance date: November 18, 2010.

The device(s)/drug(s) is/are FDA-approved or approved by corresponding national agency for this indication.

Foundation and Governmental funds were received in support of this work. No benefits in any form have been or will be received from a commercial party related directly or indirectly to the subject of this manuscript.

Gerard Urrutia and José Domingo Alarcón retrieved and independently assessed the references identified through the electronic search. They also independently assessed the methodologic quality of each article and performed data extraction. Francisco M. Kovacs participated with these two authors in solving by consensus any disagreements in quality assessment and data extraction that arose, and also wrote this manuscript. He reports that he had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Address correspondence and reprint requests to Francisco M. Kovacs, MD, PhD, Departamento Científico, Fundación Kovacs, Paseo Mallorca 36, 3º, 1a, 07012 Palma de Mallorca, Spain; E-mail: kovacs@kovacs.org

DOI: 10.1097/BRS.0b013e31820c97b1

Spine

Lumbar spinal stenosis is defined as the reduction of the surface area of the lumbar spinal canal. It is usually caused by spinal degenerative conditions^{1–5} and is commonly found to be asymptomatic.^{1–9} However, symptoms may appear as a result of neurovascular mechanisms (such as reduced arterial flow in cauda equina, venous congestion, and increased epidural pressure),^{10–13} nerve root excitation by local inflammation, or direct compression in the central canal or the lateral recess.¹⁴

These mechanisms can concur. As a result, symptoms attributed to lumbar spinal stenosis are diverse and include radicular pain down to the leg, numbness, and/or weakness. Because lumbar flexion increases the available space in the lumbar spinal canal,^{1,15} the most typical clinical feature is neurogenic claudication (pain in the buttocks or legs when walking, which disappears with sitting or lumbar flexion). However, radicular symptoms, which do not improve in flexion, can also be attributed to spinal stenosis at the corresponding level, and no clinical feature or diagnostic test is a valid diagnostic “gold standard” for concluding that spinal stenosis is the cause of pain in a given patient.¹⁶ Nevertheless, spinal

stenosis is the most common reason for lumbar spine surgery in persons older than 65 years in the United States.^{17,18}

When conservative treatments fail for 3 to 6 months, decompressive surgery is usually considered. The most recent Cochrane review concluded that there was only heterogeneous and limited evidence on its effectiveness, but it was issued in 2005.¹⁹

Therefore, the objectives of this study were to: (1) systematically review the evidence on the effectiveness and safety of any form of surgery *versus* conservative treatment for symptomatic lumbar spinal stenosis, and (2) explore whether available data made it possible to refine indication criteria for either type of treatment, on the basis of the existence of spondylolisthesis or neurogenic claudication.

MATERIALS AND METHODS

Search and Study Selection

An electronic search was performed in CENTRAL (Cochrane library 2009, issue 2), MEDLINE (January 1966–July 2009), EMBASE (January 1980–July 2009), and on the Internet through TripDatabase (only for reviews and technical reports). The terms “spinal stenosis,” “lumbar stenosis,” “claudication,” “spinal stenosis,” “surgery or surgical,” and “low back pain, lumbago, back pain or backache” were combined with the highly sensitive search strategy to identify randomized controlled trials (RCTs) developed by the Cochrane Collaboration.²⁰

Studies were included in this review if they were RCTs providing data on the comparison of the effectiveness or safety of any surgical procedure with any form of conservative treatment in patients with neurogenic claudication or sciatica, and lumbar spinal stenosis that had been confirmed by imaging. No language restriction was applied.

References identified were retrieved and assessed independently by two authors (G.U. and J.D.A.) to check for inclusion criteria. Disagreements were solved by consensus with the third author (F.M.K.).

For studies that included randomized and nonrandomized cohorts, this review considered data only from the randomized cohort to compare the effectiveness of surgical and conservative treatments, although adverse events from all patients were gathered. Separate data were also requested from the authors with regard to patients with and without spondylolisthesis, and neurogenic claudication.

Quality Assessment and Data Analysis

Methodological quality of each study was independently assessed by two of the three reviewers (G.U., J.D.A., and F.M.K.), following the criteria recommended by the Cochrane Back Review Group.^{21,22} Criteria on blindness of patients and therapists were disregarded because they were not applicable when comparing surgery with conservative treatment (Table 1). Therefore, the maximum possible score was 9 points and a study was considered as being of “high quality” if it scored 5 points or more.

Data extraction was independently undertaken (G.U. and J.D.A.). Disagreements in quality assessment and data extraction were solved by consensus (G.U., J.D.A., and F.M.K.). Despite what was planned at the design phase, neither a quantitative synthesis of data nor a sensitivity analysis was possible because of the variability of outcome measures and the heterogeneity of the methods used across the studies (Table 2). Therefore, a qualitative analysis of data was performed on the basis of the methodological quality of included trials and the consistency of their findings.

RESULTS

The electronic search identified 739 references. A total of 30 publications, corresponding to 17 studies were identified as eligible. Ten publications (nine studies) were excluded because they did not focus on spinal stenosis,^{28–37} eight publications (two studies) because they were not randomized,^{38–45} and one publication because it mixed patients who had and had not been randomized, and the authors were not able to provide their data separately.⁴⁶

The remaining 11 publications (five studies) were included in this review.^{23–27,47–52} For the purpose of this review, individual studies are identified by using the first publications reporting primary clinical endpoints.^{23–26,50} Three of these studies included data from both observational and randomized cohorts,^{23,25,26} but they were included because the authors provided separate data for those patients in the randomized cohorts.^{25,26}

All studies scored as being of high quality, although the sample size of one of them was very small.²³ Most outcomes were self-reported by the patients and blinding of care providers and patients was not feasible. Hence, outcome assessment was not blinded (Table 1).

The mean age of patients included in the studies ranged between 62 and 70 years. Imaging had documented spinal stenosis in all the participants. One study included only patients with spondylolisthesis,²⁵ one included only patients without spondylolisthesis,²⁶ one included only patients with spondylolisthesis of grade I and lesser,⁵⁰ and the other two included patients both with and without spondylolisthesis.^{23,24} When contacted, the authors from the two latter studies were unable to provide separate data for patients with and without this radiological finding (Table 2).

One study included only patients with neurogenic claudication,⁵⁰ whereas the other four included patients with pain radiating down to the leg, both with and without neurogenic claudication.^{23–26} The authors from two of these studies were able to provide separate data for patients with and without neurogenic claudication.^{25,26} Most patients included in the other two studies reported neurogenic claudication,^{23,24} although separate data on their evolution were not available (Table 2).

Four studies assessed the effectiveness of decompressive surgery (laminectomy, medial facetectomy, discectomy, and/or removal of osteophytes and hypertrophic ligament flava) with or without instrumented or noninstrumented fusion.^{23–26} The fifth study assessed the effectiveness of a specific type of interspinous implant.⁵⁰ The latter was the only study funded

TABLE 1. Quality Assessment of the Studies

	Amundsen 2005* (Randomized Cohort)	Malvimaara et al ⁴	Weinstein et al ²³ † (Randomized Cohort)	Weinstein et al ^{6,†} (Randomized Cohort)	Zucherman et al ⁷ S
1	Was the method of randomization adequate?	Yes	Yes	Yes	Yes
2	Was the treatment allocation concealed?	?	Yes	Yes	Yes
3	Were the groups similar at baseline regarding the most important prognostic factors? If they were not, were the differences taken into account at the analysis stage?	“Block randomization using tables of random numbers.” No	“Central randomization based on computer-generated random blocks of variable size, separately for each hospital.” Yes	An automated randomization system was created on the basis of computer-generated random, blocked (variable size) treatment assignments stratified by site. Yes	Patients were randomized using block randomization by a surgical center. An individual not involved in the treatment or care of the patients performed the randomization and informed the surgeon of its result. Yes
4	Were the cointerventions avoided or comparable?	The randomized cohort consisted of 31 patients (13 vs. 18) where imbalances occurred by chance. Yes	There were some imbalances (female patients, low level of education, and good perceived health were over-represented in the surgery group) between the groups. A model to control for these baseline differences between treatment groups was performed in the analysis. Yes	Details not provided (the article provides a full description of the randomized and the observational cohorts but not specifically for each treatment arm within each cohort). Nevertheless, the analysis was adjusted by all important predictors of response as covariates (all specified in the article). Yes	Same as in Weinstein et al ²⁵ . Yes
		“After discharge from the hospital after 1 mo, further treatment in both groups was identical.”		Same as in Weinstein et al ²⁵ .	Although details on the cointerventions in the surgery group are not provided, it is unlikely that this will cause bias.

(Continued)

TABLE 1. (Continued)

	Amundsen 2005* (Randomized Cohort)	Malvimaara et al ²⁴	Weinstein et al ²⁵ † (Randomized Cohort)	Weinstein et al ²⁵ † (Randomized Cohort)	Zucherman et al ²⁷ §
5	Was the compliance rate acceptable in all groups? Of the 18 patients who were randomized into the control group, 10 patients crossed over to the surgery group.	Yes Of the 50 patients who were randomized into the surgical group, 4 patients were not subjected to surgery while 4 of the 44 patients randomized into the no-surgery group were operated upon during the 2-yr follow-up.	Yes Compliance with assigned treatment was low and a significant proportion of patients in both groups crossed over to the treatment assigned for the other group. However, results were analyzed according to the "intention to treat" principle; so, this was unlikely to cause bias.	Yes Same as in Weinstein et al ²⁵ .	Yes
6	Was the care provider blinded? Not applicable	Not applicable	Not applicable	Not applicable	Not applicable
7	Was the patient blinded? Not applicable	Not applicable	Not applicable	Not applicable	Not applicable
	Was the outcome assessor blinded? ?	?	No	No	No
8	All patients were seen by the same physician (not a surgeon). Further details not provided.	Not described.	Most outcomes consisted of patient-reported outcomes.	Same as in Weinstein et al ²⁵ .	Most outcomes consisted of patient-reported outcomes.
9	Was the withdrawal/dropout rate unlikely to cause bias? Cases of 1 patient assigned to surgery and 1 patient assigned to the control group were lost to follow-up at 4 yr because of death. At 10 yr, cases of 2 patients and 1 patient, from the respective groups, were lost to follow-up for the same reason.	Yes There were 3 patients in the surgery group and 4 in the no-surgery group, the cases of whom were lost to follow-up (reasons provided).	Yes In each group, 17% of the patients were not available at the last visit (2 yr) for similar reasons. Moreover, the high rate of crossovers (42% overall) might create a bias toward the null hypothesis.	Yes In each group, 22% and 25% of the patients were not available at the last visit (2 yr) for similar reasons. Moreover, the high rate of crossovers (38% overall) might create a bias toward the null hypothesis.	? Out of 100, 9 patients assigned to control group withdrew from the study before receiving their initial epidural injection as they entered the study hoping to be randomized to surgery. Moreover, at 2-yr follow-up of the case, 7 patients in the surgery group and 10 additional patients in the control group were not available for analysis. This could eventually bias results against surgery.

(Continued)

TABLE 1. (Continued)

	Amundsen 2005* (Randomized Cohort)	Malvimaara et al ²⁴	Weinstein et al ²³ † (Randomized Cohort)	Weinstein et al ⁶⁴ ‡ (Randomized Cohort)	Zucherman et al ²⁷ §
10	Was the timing of the outcome assessment in both groups comparable?	Yes	Yes	Yes	Yes
11	Was an analysis by intention to treat performed?	Details not provided but relevant differences are unlikely to be.	Yes	Yes	No
		There were 10 patients who crossed over from control group to surgery group, forming a new group designated as group RC+. These were recorded as treatment failures in their original group on the basis of ITT analysis during the first 4 yr.	All patients were analyzed in their original group regardless of the actual treatment received. Dropouts during case follow-up were assumed to be negligible.	The primary analysis was performed according to the "intention to treat" principle. In addition, an "as treated" analysis (adjusted for potential confounding factors) was also performed to study the potential effects of crossovers.	Outcomes from patients who withdrew from the study (7 in the surgery group and 19 in the control group) were not included in the results.
		5/9	8/9	8/9	6/9
<p>Studies with multiple related publications: *Amundsen et al⁷ and Amundsen et al²³. †Birkmeyer et al⁶⁸ and Weinstein et al²⁵. ‡Birkmeyer et al⁶⁸, Weinstein et al²⁶, and Weinstein et al⁴⁹. §Zucherman et al⁵⁰, Zucherman et al²⁷, Hsu et al⁵¹, and Anderson et al⁵². ITT indicates "intention to treat"; RC, randomized cohort.</p>					

TABLE 2. Characteristics of the Included Studies

Study	Patients	Surgical Procedure	Conservative Treatment	Follow-up	Losses to Follow-up
Amundsen et al ²³ (Randomized cohort)	A total of 31 patients with sciatica together with radiologic signs of stenosis and compression of the clinically afflicted nerve roots. Disc herniations and spondylolysis excluded. Most patients with neurogenic claudication (number unknown). No. of patients with spondylolisthesis unknown.*	n = 13 patients. Decompression (laminectomy, facetectomy, discectomy, removal of osteophytes, and hypertrophic ligament flava) without fusion. + Stabilizing orthosis for 3 mo, "back school," rehabilitation for 1 mo, and subsequent stabilizing exercises.	n = 18 patients. Stabilizing orthosis (for all activities during the day) for 3 mo, "back school," rehabilitation for 1 mo, and subsequent stabilizing exercises.	6 mo, 1, 4, and 10 yr	Surgery: 1 (4 yr), 2 (10 yr) Control: 2 (4 yr), 4 (10 yr)
Malmivaara et al ⁴	A total of 94 patients (mean age: 62 yr) with pain irradiating to legs or buttocks for ≥6 mo, fatigue or loss of sensation in the legs aggravated by walking, no neurologic dysfunction, and spinal stenosis (sagittal diameter of the dural sac <10 mm). No. of patients with spondylolisthesis or neurogenic claudication unknown.†	n = 50 assigned (45 were actually operated on). Segmental decompression and undercutting facetectomy, with or without fusion (with or without transpedicular instrumentation). + Brochure on spinal stenosis and the principles of activation and physical training, education on pain-relieving body postures, ergonomics, and individualized exercises.	n = 44 assigned (40 actually treated without surgery). NSAIDs, physiotherapy (ultrasound and exercise) for 24% of the patients. + Brochure on spinal stenosis and the principles of activation and physical training, education on pain-relieving body postures, ergonomics, and individualized exercises.	6 mo, 1 and 2 yr	Surgery: 3 (2 yr) Control: 4 (2 yr)
Weinstein et al ²⁵ (Randomized cohort)	A total of 304 patients (mean age: 66 yr) with neurogenic claudication or sciatica for ≥3 mo, and spinal stenosis and degenerative spondylolisthesis seen on imaging. Spondylolysis and isthmia spondylolisthesis excluded. Spondylolisthesis in 100%. Neurogenic claudication in 86%.	n = 159 assigned (101 were actually operated on). Posterior decompressive laminectomy with or without bilateral single-level fusion (with or without posterior pedicle screw instrumentation).	n = 145 assigned (70 were actually treated without surgery). At least active physical therapy, education, or counseling including instructions for exercising at home, NSAIDs if tolerated.	6 wk, 3 and 6 mo, 1, 2, and 4 yr [‡]	Surgery: 15 (1 yr), 27 (2 yr), 52 (4 yr) Control: 11 (1 yr), 25 (2 yr), 46 (4 yr)
Weinstein et al ²⁶ (Randomized cohort)	A total of 289 patients (mean age: 65.5 yr) with neurogenic claudication or sciatica for ≥3 mo and spinal stenosis seen on imaging. Spondylolysis and spondylolisthesis, excluded. Spondylolisthesis in 0%. Neurogenic claudication in 79%.	n = 138 assigned (92 were actually operated on). Posterior decompressive laminectomy.	n = 151 assigned (86 were actually treated without surgery). At least active physical therapy, education, or counseling including instructions for exercising at home, NSAIDs if tolerated.	6 wk, 3 and 6 mo, 1, 2, and 4 yr [‡]	Surgery: 18 (1 yr), 30 (2 yr), 54 (4 yr) Control: 25 (1 yr), 38 (2 yr), 65 (4 yr)

(Continued)

TABLE 2. (Continued)

Study	Patients	Surgical Procedure	Conservative Treatment	Follow-up	Losses to Follow-up
Zucherman et al ²⁷ (2005)	A total of 200 patients, ≥50 yr (mean age: 70 yr), able to walk ≥50 feet, with leg, buttock, or groin pain that was relieved during flexion. Motor deficit, spondylolisthesis > Grade I, and previous lumbar surgery excluded. Spondylolisthesis grade I in 39%. [†] Neurogenic claudication in 100%.	n = 100 Interspinous Process Distraction System (X STOP).	n = 100 (91 were actually treated without surgery). Patients randomized to the control group received at least one epidural steroid injection after enrollment and were prescribed additional epidural steroid injections, nonsteroidal anti-inflammatory medications, analgesics, and physical therapy as necessary. Physical therapy consisted of back school and methods such as ice packs, heat packs, massage, stabilization exercises, and pool therapy.	6 wk, 6 mo, 1 and 2 yr	Surgery: 7 (2 yr) Control: 19 (2 yr)

*A total of 100 patients were recruited in this study, 91 of whom had neurogenic claudication. Among those 100 patients, only 31 were randomized. Mean age of those 31 patients and how many of them had spondylolisthesis or neurogenic claudication is unknown. These data are not given in the article and the authors could not provide the same.

[†]Anderson et al²⁸.

[‡]Weinstein et al⁴⁹.

[§]Data supplied by the authors of the original study.

by the industry, as opposed to public or not-for-profit sources, in which at least one author had potentially relevant conflict of interest.⁵⁰

Treatments in the conservative arms of those studies included different or nonspecified types of orthosis, rehabilitation, physical therapy, exercise, nonsteroidal anti-inflammatory drugs, analgesics, and education, as well as heat and cold, transcutaneous electrical nerve stimulation, ultrasound, and epidural steroids.

The decision on the kind of surgery or conservative treatment applied to each patient was left up to the therapists and no explicit criteria were used.

Follow-up ranged between 2 and 10 years. Variables that were assessed during follow-up included pain, function, claudication distance, quality of life, and patients' overall rating and satisfaction, although instruments used to assess those variables varied across studies (Table 3). Disability, as measured by the Oswestry Disability Index at 2 years, was the only outcome that was assessed in 3 studies or more (Figure 1).²⁴⁻²⁶ Losses to follow-up ranged between 7.7% (at 1 year) and 44.0% (at 4 years), and were roughly similar between the surgical and conservative arms (Table 2).

Small improvements were generally seen in the patients treated conservatively and in whom no catastrophic complications such as cauda equina syndrome arose. Changes in claudication distance were similar between conservative and surgical groups.²⁴ However, results of all the studies consistently favored decompressive surgery (with or without fusion) for improvement of pain, function, and quality of life, as well as in terms of patients' satisfaction.²³⁻²⁶ Crossover rates ranged between 9% and 54%, and affected patients assigned to both the conservative and surgical groups (Table 3). In the two studies with higher crossover rates,^{25,26} differences between conservative and surgical treatments were statistically significant only in the "as treated" analysis. The only exception was 36-item Medical Outcomes Study Short-Form General Health Survey (SF-36)—"bodily pain," which even in the "intention to treat" analysis favored surgery at 2 years in one of these studies.²⁶

The advantage of surgery was already noticeable at 3 to 6 months and remained constant for up to 2 to 4 years, although by the end of that period the differences tended to be smaller (Table 3). In general, the magnitude of the additional improvement triggered by surgery in pain and function was clinically relevant.^{53,54} Surgery was more effective than conservative treatment for patients with and without spondylolisthesis. Surgery was also more effective than conservative treatment for patients with and without neurogenic claudication, although the advantage of surgery was generally greater for the former (Table 3).

The interspinous implant also led to a better evolution than conservative treatment with regard to pain, function, quality of life, and patients' satisfaction.^{27,50-52} Differences were statistically significant at 6 weeks and remained so till 2 years later. Results favored the implant over conservative treatment to the same extent in patients with and without (Grade I) spondylolisthesis. This radiological feature was not predictive of outcome.⁵²

Adverse events among patients treated conservatively included injection intolerance, symptom flare, leg paresthesia, and increased back pain. Among patients who underwent decompressive surgery, with or without spinal fusion, between 5.4% and 14% suffered from perioperative complications (the most common being, in all studies, dural tears). Postoperative complications arose in between 8.2% and 18% of the patients and included pulmonary edema, peridural hematoma, sepsis, and misjudgment of stenotic level. Reported reoperation rates were 1.3% to 2% at 1 year, 6% to 11% at 2 years, and 15% at 4 years. The rate of recurrent stenosis at 4 years was 5%. The mean duration of the operation was 120 to 199 minutes and the mean blood loss was 198 to 589 mL.^{25,26} The number of deaths were similar in the surgical and conservative groups. The overall postoperative death rate ranged between 0.3% and 0.5%. Approximately 90% of the patients undergoing surgery did not suffer from any adverse event.

Adverse events related to the implantation of the interspinous distraction device included asymptomatic spinous process fracture (1%), malpositioned implant (1%), implant dislodgement or migration (1%), increased pain at implant level (1%), wound dehiscence (1%), wound swelling (1%), incisional pain (1%), and hematoma (1%). In addition, one patient suffered from respiratory distress, another from a transient ischemic coronary episode, and a patient with a history of cardiovascular disease died from pulmonary edema 2 days after device implantation. The reported mean blood loss was 46 mL, the mean duration of the procedure was 54 minutes, and 96% of the patients stayed in the hospital for less than 24 hours. Among the patients in whom the interspinous distraction device was implanted, 7% subsequently underwent a laminectomy. Among those with spondylolisthesis, this percentage was 11.9%.

DISCUSSION

These results suggest that in patients with radicular pain caused by spinal stenosis, in whom conservative treatment has failed for 3 to 6 months, decompressive surgery (with or without fusion) does not improve walking ability but improves pain, function, and quality of life to a greater extent than continuing conservative treatment. Surgery is superior to conservative treatment irrespective of the patients' degree of affectation whether they have spondylolisthesis or neurogenic claudication. However, slightly better results for surgery can be anticipated in patients with neurogenic claudication. Patients should be aware that the benefits of surgery decrease with time but still remain significant till 4 (and possibly 10) years later.^{23,49} Results in the longer term are unknown. However, obtaining a clinically relevant improvement for such a period is worthwhile even if it were to fade thereafter.

Surgery for spinal stenosis can be effective despite advanced age, multilevel involvement, or common comorbidities such as diabetes, obesity, chronic coronary disease, or chronic lung diseases.^{26,55-57} However, these features, especially the latter, and hospitalizations in the previous year are associated with more complications and mortality.⁵⁵ Therefore, the risks of

surgery should be balanced against expected improvements for each individual patient.

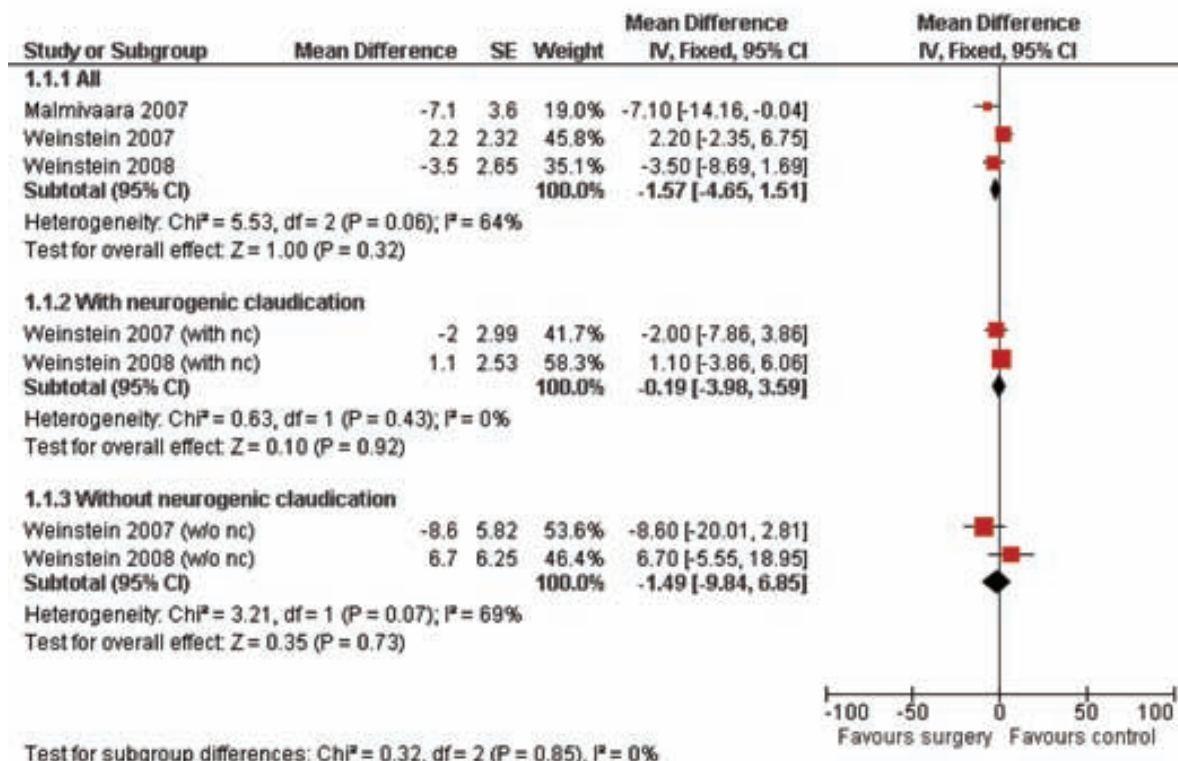
In patients aged 50 years or older with neurogenic claudication and spondylolisthesis of Grade I who are able to walk 50 feet or more and have no motor deficit, the implantation of a specific type of interspinous distraction device also leads to better results than conservative treatment for up to 2 years.²⁷ Results beyond this time period are unknown. The evidence supporting these results derives from industry-funded studies whose authors harbored conflicts of interests.²⁷ The effectiveness or safety of other interspinous implants is unknown and no study has directly compared decompressive surgery with the interspinous implant.

Only among patients who have spondylolisthesis, adding fusion to decompression may improve the results of surgery for lumbar spinal stenosis.⁵⁸⁻⁶⁰ However, fusion increases major complications, 30-day mortality, and resources used.⁵⁵ In patients who have fusion, instrumented and noninstrumented fusion lead to similar results.^{27,61,62} Among patients who undergo instrumented fusion, simpler procedures (*i.e.*, anterior or transverse process or posterior fusion, at only 1 or 2 levels) are safer than complex ones (*i.e.*, any combination of simple procedures, 360° fusion, or fusion of more than two levels)^{19,55,62-64} and lead to similar results.⁶¹

In all the studies included in this review, each care provider decided the form of conservative or surgical treatment applied to a given patient without following any predefined indication criteria. This approach was probably the only feasible one, but it led to treatments within the conservative and surgical groups being as heterogeneous as they are in routine practice both across studies and across patients in the same study. This precludes any conclusion on the comparative effectiveness of any particular form of surgery or conservative treatment. However, surgery was more effective than conservative treatment across all the studies and results of surgery were roughly consistent among different surgeons in trials conducted in the United States and Scandinavia. This may suggest that their criteria were somehow consistent or that potential differences in their criteria or in the different surgical techniques used do not have a significant influence on results.

Data from observational or nonrandomized cohorts were excluded from this review because of concerns regarding the risk of unknown biases. However, despite being classified as "high quality" according to methodological criteria,^{21,22} concerns exist with regard to the studies included in this review. One study recruited only 31 patients,²³ another included only 94,²⁴ and none conducted a power analysis. As a result, sample sizes may have been insufficient to detect differences for some outcomes. Treatments used in both the surgical and conservative groups were heterogeneous and were not selected according to explicit criteria. To be eligible, patients had to have previously undergone conservative treatment unsuccessfully, which may have made it difficult to recruit large samples, affected the comparability of the unspecific effects across groups (*e.g.*, Hawthorne or placebo), and increased rates of withdrawal, dropout, and crossover.^{23,25,26} Although results consistently favored surgery, in the two studies with

(A) ODI at 2 yrs (ITT analysis)



(B) ODI at 2 yrs (As treated)

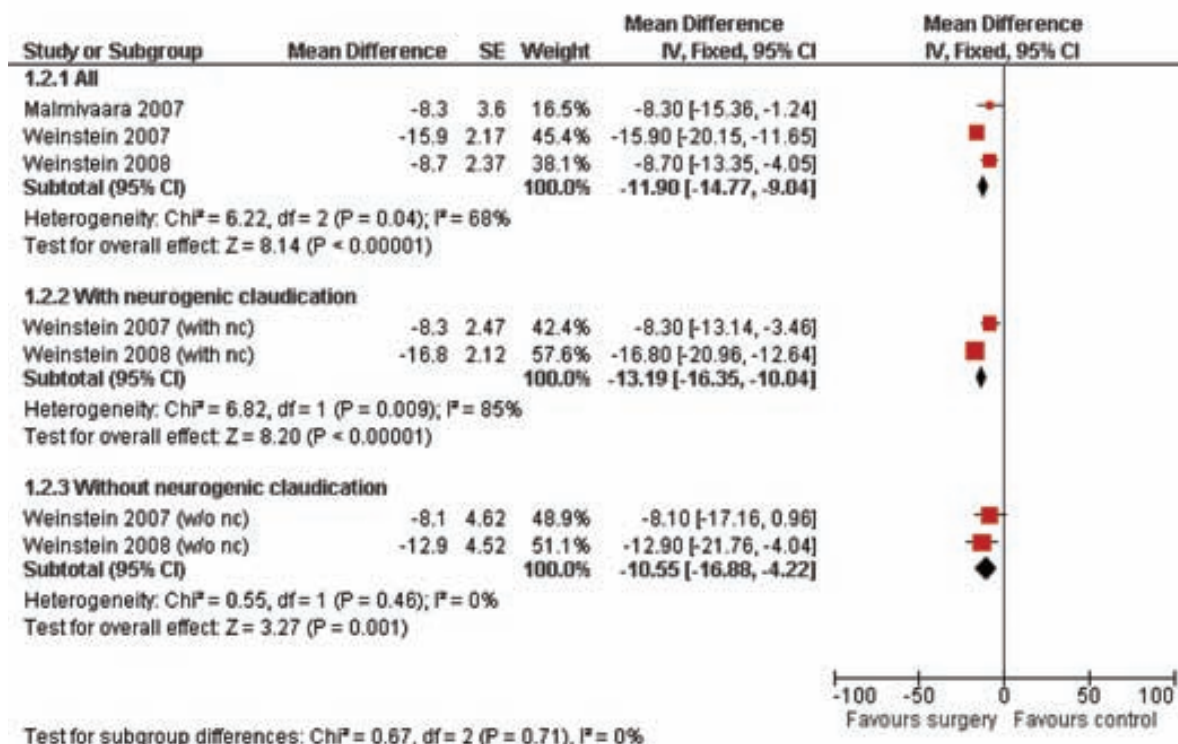


Figure 1. Forest plot showing the results on disability in the three studies that used this instrument, according to the (A) “intention to treat” and (B) “as treated.” In these studies, the disability was measured by the Oswestry Disability Index at 2 years.

TABLE 3. Summary of Results

Study	Patient's Global Assessment* ("Good Result") [†]	Pain ("Without Pain" or "Mild Pain") [‡]	Leg Pain [§]	Back Pain [§]	Disability (ODI) [†]	SF-36 Bodily Pain	SF-36 Physical Function ^{**}	Other Outcomes
Amundsen et al ³ (Randomized cohort)	1 yr: RR 2.07 (0.98–4.38) 4 yr: RR 1.94 (1.14–3.31) 10 yr: RR 3.18 (0.97–10.41)	1 yr: NR 4 yr: RR 3.33 (0.77–14.33)	1 yr: MD 1.69 (0.41–2.96) 2 yr: MD 1.51 (0.25–2.77)	1 yr: MD 2.33 (1.12–3.55) 2 yr: MD 2.13 (0.98–3.28)	1 yr: MD 11.3 (4.3–18.8) 2 yr: MD 7.8 (0.8–14.9) ≥ 10-point reduction (ODI): 1 yr: RR 2.16 (1.31–3.57) 2 yr: RR 1.36 (0.88–2.10)			Other outcomes (claudication or walking distance; level of daily activity; and neurologic deficits) were not reported separately for the randomized cohort.
Malmivaara et al ⁴								Walking disability (walking distance <1,250 m) ^{***} 1 yr: RR 0.93 (0.61–2.03) 2 yr: RR 1.08 (0.70–2.42) Walking disability (walking distance <400 m) ^{***} 1 yr: RR 0.91 (0.51–4.24) 2 yr: RR 1.18 (0.67–4.72)
Weinstein et al ⁵ (Randomized cohort—all)					"Intention to treat": 2 yr: DMC 2.2 (–2.3 to 6.8) 4 yr: DMC 4.1 (–0.8 to 9.1) "As treated": 2 yr: DMC –15.9 (–20.2 to –11.7) 4 yr: DMC –16.2 (–20.7 to –11.6)	"Intention to treat": 2 yr: DMC 1.5 (–4.2 to 7.3) 4 yr: DMC –2 (–8.6 to 4.6) "As treated": 2 yr: 17.8 (12.5 to 23) 4 yr: 17.1 (10.9 to 23.4)	"Intention to treat": 2 yr: DMC 1.9 (–3.7 to 7.5) 4 yr: DMC –3.1 (–9.2 to 3.0) "As treated": 2 yr: 16.7 (11.4 to 22.1) 4 yr: DMC 19.2 (13.4 to 25.1)	Other outcomes (patient's satisfaction; Stenosis Bothersomeness Index, Leg Pain Bothersomeness Scale; and Low Back Pain Bothersomeness Scale) were not provided separately for the randomized cohort.
(Randomized cohort—with neurogenic claudication)					"Intention to treat": 2 yr: DMC –2 (–7.8 to 3.9) 4 yr: DMC –1.2 (–7.5 to 5.1) "As treated": 2 yr: DMC –8.3 (–13.1 to –3.4) 4 yr: DMC –6.1 (–12.3 to 0.1)	"Intention to treat": 2 yr: DMC 5.9 (–1.8 to 13.5) 4 yr: DMC 0.1 (–8.3 to 8.4) "As treated": 2 yr: DMC 11.3 (5.1 to 17.6) 4 yr: DMC 8.8 (0.6 to 16.9)	"Intention to treat": 2 yr: DMC –3.4 (–10.6 to 3.9) 4 yr: DMC –3.1 (–10.9 to 4.8) "As treated": 2 yr: DMC 5.1 (–0.8 to 11.1) 4 yr: DMC 6.9 (–0.7 to 14.5)	

(Continued)

TABLE 3. (Continued)

Study	Patient's Global Assessment* ("Good Result")†	Pain ("Without Pain" or "Mild Pain")‡	Leg Pain§	Back Pain§	Disability (ODI)†	SF-36 Bodily Pain	SF-36 Physical Function**	Other Outcomes
(Randomized cohort—without neurogenic claudication)					"Intention to treat": 2 yr: DMC -8.6 (-20 to 2.8) 4 yr: DMC 4.2 (-8.1 to 16.5) "As treated": 2 yr: DMC -8.1 (-17.2 to 0.9) 4 yr: DMC -8.8 (-20.2 to 2.7)	"Intention to treat": 2 yr: DMC 14.9 (0.1 to 29.7) 4 yr: DMC 2 (-13.9 to 17.9) "As treated": 2 yr: DMC 13.7 (2.3 to 25.2) 4 yr: DMC 18.7 (4.5 to 33)	"Intention to treat": 2 yr: DMC 11.6 (-2.4 to 25.6) 4 yr: DMC -4.5 (-19.5 to 10.5) "As treated": 2 yr: DMC 11.4 (0.5 to 22.2) 4 yr: DMC 11.8 (-1.6 to 25.3)	
Weinstein et al ⁶ (Randomized cohort—all)					"Intention to treat": 2 yr: DMC -3.5 (-8.7 to 1.7) "As treated": 2 yr: -8.7 (-13.3 to -4.0)	"Intention to treat": 2 yr: DMC 7.8 (1.5 to 14.1) "As treated": 2 yr: 11.7 (6.2 to 17.2)	"Intention to treat": 2 yr: DMC 0.1 (-6.4 to 6.5) "As treated": 2 yr: 8.1 (2.8 to 13.5)	Same as in Weinstein et al ⁶
(Randomized cohort—with neurogenic claudication)					"Intention to treat": 2 yr: DMC +1.1 (-3.9 to 6) 4 yr: DMC+ -3.7 (-1.6 to 9.0) "As treated": 2 yr: DMC -16.8 (-20.9 to -12.6) 4 yr: DMC -18.4 (-23.3 to -13.5)	"Intention to treat": 2 yr: DMC 2 (-4.6 to 8.7) 4 yr: DMC -3.6 (-10.8 to 3.5) "As treated": 2 yr: DMC 20 (14.3 to 25.6) 4 yr: DMC 20.6 (13.9 to 27.3)	"Intention to treat": 2 yr: DMC 3.2 (-2.9 to 9.2) 4 yr: DMC -3 (-9.6 to 3.5) "As treated": 2 yr: DMC 17.9 (12.6 to 23.3) 4 yr: DMC 20.4 (14.1 to 26.8)	
(Randomized cohort—with neurogenic claudication)					"Intention to treat":	"Intention to treat":	"Intention to treat":	

(Continued)

TABLE 3. (Continued)

Study	Patient's Global Assessment* ("Good Result")†	Pain ("Without Pain" or "Mild Pain")‡	Leg Pain§	Back Pain§	Disability (ODI)†	SF-36 Bodily Pain¶	SF-36 Physical Function**	Other Outcomes
Zucherman et al ²⁷ 2005	2 yr: 73.1% (surgery) vs. 35.9% (control) (P < 0.001)				2 yr: DMC 6.7 (-5.6 to 18.9) 4 yr: DMC 5.7 (-7.6 to 19.0) "As treated": 2 yr: DMC -12.9 (-21.8 to -4.1) 4 yr: DMC -10.2 (-21.2 to 0.8)	2 yr: DMC -2.3 (-18.5 to 13.9) 4 yr: DMC 9.3 (-8.7 to 27.4) "As treated": 2 yr: DMC 16.3 (4.5 to 28.1) 4 yr: DMC 1.9 (-13.2 to 17)	2 yr: DMC -1.9 (-16.7 to 12.8) 4 yr: DMC -2.8 (-19.3 to 13.7) As treated": 2 yr: DMC 8.1 (-3.1 to 19.2) 4 yr: DMC 18.9 (4.7 to 33.1)	Symptoms Severity Score†† Surgery better at 6 wk, 6 mo, 1 and 2 yr (graphs) (P < 0.001) 2 yr: MPC 45.4% (surgery) vs. 7.4% (control) (P < 0.001) "Clinically relevant improvement (patients)"/: 2 yr: 60.2% (surgery) vs. 18.5% (control) (P < 0.001) Symptoms Severity Score†† Surgery better at 6 wk, 6 mo, 1 and 2 yr (graphs) (P < 0.001) 2 yr: MPC 44.3% (surgery) vs. -0.4% (control) (P < 0.001) "Clinically relevant improvement (as measured by patients)"/: 2 yr: 57% (surgery) vs. 14.8% (control) (P < 0.001)

(Continued)

TABLE 3. (Continued)

Study	Patient's Global Assessment* ("Good Result")†	Pain ("Without Pain" or "Mild Pain")‡	Leg Pain§	Back Pain§	Disability (ODI)†	SF-36 Bodily Pain	SF-36 Physical Function**	Other Outcomes
(Subgroup with spondylolisthesis)								ZCQ ^{§§} (global success) ^{¶¶} 6 mo: 52% (surgery) vs. 9% (control) (<i>P</i> value not reported) 1 yr: 59% vs. 12% (<i>P</i> value not reported) 2 yr: 48.4% (surgery) vs. 4.9% (control) (<i>P</i> < 0.0001) Quality of life (SF-36) At all post-treatment time points (6 wk, 6 mo, 1 yr, and 2 yr), the mean domain scores documented in the X STOP group were significantly greater than those in the nonoperative group, with the exception of the mean General Health, Role Emotional, and Mental Component Summary scores at 2 yr
	Patient's satisfaction (ZCQ domain); ^{##} 2 yr: Better surgery (statistically significant) (graphs)							Symptoms Severity Score ^{††} 2 yr: Better surgery (<i>P</i> < 0.0001) Symptoms Severity Score ^{††} 2 yr: Better surgery (<i>P</i> < 0.0001) ZCQ ^{§§} (global success) ^{¶¶} 2 yr: 63.4% (surgery) vs. 12.9% (control) ("highly statistically significant")

(Continued)

TABLE 3. (Continued)

Study	Patient's Global Assessment* ("Good Result") [†]	Pain ("Without Pain" or "Mild Pain") [‡]	Leg Pain [§]	Back Pain [§]	Disability (ODI) [†]	SF-36 Bodily Pain	SF-36 Physical Function ^{**}	Other Outcomes
<p>*The patients were asked to evaluate their own situation and to state whether it was better, worse, or unchanged as compared with their condition on entering the study. This, together with the opinion of the examining physician, constituted the basis for the statement of overall treatment result (excellent, fair, unchanged, and worse).</p> <p>†Excellent (full or almost full restitution considering pain and physical function) or fair (partial restitution with lesser problems and clear improvement over the condition at entry to the study) was considered to be a good result.</p> <p>‡The intensity of pain was stated verbally (severe, moderate, or light).</p> <p>§The intensity of pain was assessed by a Numerical Rating Scale (0, no pain to 10, worst possible pain).</p> <p>¶ Oswestry Disability Index, where 0 indicates no disability and 100 indicates worst possible disability. Weinstein, 2007 and 2008, expressed the results as the percentage of the maximum score.</p> <p> SF-36 Health Status Questionnaire (physical function subscale), where higher scores indicate better outcomes.</p> <p>** SF-36 Health Status Questionnaire (pain subscale), where higher scores indicate better outcomes.</p> <p>†† A domain of the ZCQ (1–4), where 1 is the best possible outcome and represents no pain.</p> <p>††† A domain of the ZCQ (1–5), where 1 is the best possible outcome and represents no limitation in function.</p> <p>§§ Zurich Claudication Questionnaire (captures data in three distinct domains: symptom severity; physical function; and post-treatment patient satisfaction) (3–13), where 3 is the best possible outcome.</p> <p>¶¶ Treatment is considered successful if the patient is at least "somewhat satisfied" (an average score of 2.5) and has at least a 0.5 improvement in both symptom severity and physical function.</p> <p> Data provided by the authors at request (not published).</p> <p>*** Walking capacity was assessed on a treadmill at a speed of 2.5 km/h for a maximum of 30 min or 1.25 km. Restriction of walking was defined as the distance after which a patient did not want to go on because of pain, or fatigue or weakness in legs, or a combination of these.</p> <p>36-item Medical Outcomes Study Short-Form General Health Survey</p> <p>DMC indicates differences in mean changes from baseline (with 95% confidence interval); MD, mean difference (with 95% confidence interval); MPC, mean percentage change (with 95% confidence interval); RR, risk ratio (with 95% confidence interval).</p> <p>Bold letters indicate the results that favor surgery (significant differences).</p>								

the largest crossover rates, most differences were statistically significant only in the “as treated” analysis and not in the “intention to treat” one.^{25,26} Moreover, blindness (of therapists, patients, and outcome assessors) is unfeasible in studies comparing surgical with nonsurgical procedures and the placebo effect from surgery is likely to be more powerful than the one from conservative treatment, especially among patients in whom the latter has already failed. This is a special concern because outcome variables are subjective and assessed by the patients (e.g., pain severity, function, and satisfaction).

However, these are the variables that make clinical sense, were assessed using previously validated methods, the studies’ design was reasonable (Table 1), and their limitations are inherent to the kind of treatments being assessed and to the patient characteristics. The only way of controlling for potential differences in the placebo effect would be to undertake RCTs comparing decompressive surgery with sham surgery, which might face ethical and practical difficulties. Despite differences in methods and geographical settings, results from the studies included in this review consistently favor surgery and are biologically plausible. All this suggests that, although methodological weaknesses make it inappropriate to consider each of the reviewed studies as representing “the best evidence in ideal conditions,” the general conclusions from this review may be seen as deriving from “the best possible evidence in this field.”

Further RCTs could also compare different surgical techniques in specific subsets of patients and should compare the interspinous device with surgery in the subset of patients in whom the former is indicated. Such a study appears to have been initiated.⁶⁵ Further RCTs could also assess the effectiveness of specific forms or combinations of conservative treatments to compare surgery with truly effective alternatives. All these studies should be of high methodological quality and conducted independently from the industry by authors with no conflicts of interest.⁶⁶ They should be designed as large multicenter trials with strategies to reduce withdrawals and crossovers. There may be ethical concerns regarding further RCTs using the same methods as those used in the studies included in this review to compare the same miscellanea of surgical and conservative treatments.

No catastrophic events (e.g., cauda equina syndrome) arose among the patients receiving conservative treatment. Results from surgery were similar among patients initially assigned to surgery and those who crossed over to this group after failure of the conservative treatment they had been initially assigned to.^{23,25,26} This suggests that delaying surgery is not associated with a poorer prognosis. On the contrary, the interspinous implant is less aggressive and resulted in less blood loss and a shorter hospital stay than decompressive surgery (with or without fusion). Taken together, these results may counsel a step-by-step approach in the treatment of spinal stenosis. The first step could be a conservative treatment for 3 to 6 months. If such an approach fails, the implantation of a specific type of interspinous device should be considered in patients aged 50 years and older, with neurogenic claudication and spondylolisthesis of Grade I or less who are able to walk 50 feet or more and have no motor deficit. If these indication criteria

are not present, decompressive surgery could be envisaged. In patients with significant spondylolisthesis, the surgeon may consider complementing decompressive surgery with noninstrumented posterolateral fusion or with the simplest possible instrumented fusion procedure. This approach would not be suitable for patients who show spinal stenosis on imaging but are asymptomatic or report only back pain (without pain or neurological symptoms in the legs), or for those with serious neurological impairment, especially in those with cauda equina syndrome in whom surgery should be conducted urgently.⁶⁷

Generalizability of these results to clinical practice should be discussed. Care providers participating in these RCTs were volunteers as opposed to a random selection of those working in the setting where the studies took place and whether their criteria and level of skills are representative of the rest of care providers in their environment is unknown. Moreover, no data are available on the learning curves necessary before application of some of the forms of treatment that were used in these studies, such as epidural injections or surgery. Education and training standards for these procedures in routine practice should match those required in these RCTs because these standards are essential for generalizability of any nonpharmacological health technology.⁶⁶ Postmarketing surveillance should also be implemented, ideally by setting up a registry including all patients seeking care for symptomatic spinal stenosis, minimizing losses to follow-up, and using validated methods to gather clinically relevant data (including demographic data, clinical features, comorbidities, conservative and surgical treatments applied to each patient, experience and training standards of the care providers applying them, and patient’s clinical evolution).⁶⁶ Such postmarketing surveillance would not substitute RCTs but would be valuable to: (1) determine the effects in routine clinical practice, of treatments that have previously shown to be effective through RCTs, (2) compare, in routine practice conditions, the results from treatments that have shown to be effective for the same subset of patients as in previous RCTs, (3) identify factors that predict the result of each treatment to refine its indication criteria, and (4) gather data on unexpected safety concerns.

In conclusion, this review shows that decompressive surgery with or without fusion and implantation of a specific interspinous distraction device are more effective than continued conservative treatment for radicular pain due to spinal stenosis in patients in whom the latter has failed for 3 to 6 months.

➤ Key Points

- ❑ Five high-quality randomized controlled trials (RCTs), including a total of 918 patients, compared surgical procedures (the implantation of a specific type of interspinous distraction device or decompressive surgery, with or without fusion, whether instrumented or not) with a miscellanea of conservative treatments, in patients with symptomatic lumbar spinal stenosis in whom the latter had failed for 3 to 6 months.

- ❑ In all the studies, surgery led to better results for pain, disability, and quality of life, although not for walking ability. Results of surgery were similar among patients with and without spondylolisthesis, and slightly better among those with neurogenic claudication than among those without it. The advantage of surgery was noticeable at 3 to 6 months and remained for up to 2 to 4 years, although at the end of that period differences tended to be smaller.
- ❑ Main methodological concerns in these RCTs include the heterogeneity of treatments within the conservative and surgical groups, the fact that all patients included had unsuccessfully undergone conservative treatments previously, which may have led to differences in the placebo effect across groups and increased crossover, withdrawal, and dropout rates, and the lack of blindness of care provider, patients, and outcome assessment. However, these concerns are inherent to the type of patients included and the kind of treatments compared. Therefore, evidence from these RCTs can be seen as the "best possible" in this field. There may be ethical concerns regarding further RCTs comparing the forms of surgery and conservative treatments, which have been used in these trials.
- ❑ Further RCTs should: (1) compare the interspinous device with surgery in patients aged 50 years or older, with neurogenic claudication and spondylolisthesis of Grade I and lesser, who are able to walk 50 feet or more, and have no motor deficit; (2) compare different surgical techniques and interspinous devices in specific subsets of patients; (3) define indication criteria for the different surgical procedures; (4) define learning curves for all these procedures; (5) implement registries and postmarketing surveillance methods; and (6) not be conducted by the industry or by authors who harbor conflicts of interest.

References

1. Epstein NE, Maldonado VC, Cusick JF. Symptomatic lumbar spinal stenosis. *Surg Neurol* 1998;50:3-10.
2. Alvarez JA, Hardy RH Jr. Lumbar spine stenosis: a common cause of back and leg pain. *Am Fam Physician* 1998;57:1825-40.
3. Postacchini F. The diagnosis of lumbar spinal stenosis: analysis of clinical and radiographic findings in 43 cases. *Ital J Orthop Traumatol* 1985;11:5-21.
4. Tan SB. Spinal canal stenosis. *Singapore Med J* 2003;44:168-9.
5. Jacobsen S, Sonne-Holm S, Roving H, et al. Degenerative spondylolisthesis: an epidemiological perspective: the Copenhagen Osteoarthritis Study. *Spine* 2007;32:120-5.
6. Boden SD, Davis DO, Dina TS, et al. Abnormal magnetic-resonance scans of the lumbar spine in asymptomatic subjects: a prospective investigation. *J Bone Joint Surg Am* 1990;72:403-8.
7. Haig AJ, Tong HC, Yamakawa KSJ, et al. Predictors of pain and function in persons with spinal stenosis, low back pain, and no back pain. *Spine* 2006;31:2950-7.
8. Kalichman L, Li L, Kim HD, et al. Facet joint osteoarthritis and low back pain in the community-based population. *Spine* 2008;33:2560-5.
9. Kalichman L, Kim HD, Li L, et al. Spondylolysis and spondylolisthesis prevalence and association with low back pain in the adult community-based population. *Spine* 2009;34:199-205.
10. Kobayashi S, Baba H, Uchida K, et al. Blood circulation of cauda equina and nerve root [in Japanese]. *Clin Calcium* 2005;15:63-72.
11. Porter RW. Spinal stenosis and neurogenic claudication. *Spine* 1996;21:2046-52.
12. Takahashi K, Kagechika K, Takino T, et al. Changes in epidural pressure during walking in patients with lumbar spinal stenosis. *Spine* 1995;20:2746-9.
13. Takahashi K, Miyazaki T, Takino T, et al. Epidural pressure measurements. Relationship between epidural pressure and posture in patients with lumbar spinal stenosis. *Spine* 1995;20:650-3.
14. Kobayashi S, Kokubo Y, Uchida K, et al. Effect of lumbar nerve root compression on primary sensory neurons and their central branches: changes in the nociceptive neuropeptides substance P and somatostatin. *Spine* 2005;30:276-82.
15. Hall S, Bartleson JD, Onofrio BM, et al. Lumbar spinal stenosis: clinical features, diagnostic procedures, and results of surgical treatment in 68 patients. *Ann Intern Med* 1985;103:271-5.
16. DeGraaf I, Prak A, Bierma-Zeinstra S, et al. Diagnosis of lumbar spinal stenosis. A systematic review of the accuracy of diagnostic tests. *Spine* 2006;31:1168-76.
17. Deyo RA, Ciol MA, Cherkin DC, et al. Lumbar spinal fusion: a cohort study of complications, reoperations, and resource use in the Medicare population. *Spine* 1993;18:1463-70.
18. Deyo RA, Gray DT, Kreuter W, et al. United States trends in lumbar fusion surgery for degenerative conditions. *Spine* 2005;30:1441-5.
19. Gibson JN, Waddell G. Surgery for degenerative lumbar spondylosis; updated cochrane review. *Cochrane Database Syst Rev* 2005;19(4):CD001352.
20. Higgins JPT, Green S, eds. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.2*. Oxfordshire: The Cochrane Collaboration, 2008. Updated September 2008. Available at: <http://www.cochrane-handbook.org>. Accessed May 17, 2010.
21. Van Tulder MW, Assendelft WJJ, Koes BW, et al. Method guidelines for systematic reviews in the Cochrane Collaboration Back Review Group for spinal disorders. *Spine* 1997;22(20):2323-30.
22. Van Tulder M, Furlan A, Bombardier C, et al. Updated method guidelines for systematic reviews in the Cochrane Collaboration Back Review Group. *Spine* 2003;28:1290-9.
23. Amundsen T, Weber H, Nordal HJ, et al. Lumbar spinal stenosis: conservative or surgical management? A prospective 10-year study. *Spine* 2000;25:1424-36.
24. Malmivaara A, Slati P, Heliövaara M, et al. Surgical or nonoperative treatment for lumbar spinal stenosis? A randomized controlled trial. *Spine* 2007;32:1-8.
25. Weinstein JN, Lurie JD, Tosteson TD, et al. Surgical versus nonsurgical treatment for lumbar degenerative spondylolisthesis. *N Engl J Med* 2007;356(22):2257-70.
26. Weinstein JN, Tosteson TD, Lurie JD, et al. SPORT Investigators. Surgical versus nonsurgical therapy for lumbar spinal stenosis. *N Engl J Med* 2008;358:794-810.
27. Zucherman JF, Hsu KY, Hartjen CA, et al. A multicenter, prospective, randomized trial evaluating the X STOP interspinous process decompression system for the treatment of neurogenic intermittent claudication: two-year follow-up results. *Spine* 2005;30(12):1351-8.
28. Osterman H, Seitsalo S, Karppinen J, et al. Effectiveness of microdiscectomy for lumbar disc herniation: a randomized controlled trial with 2 years of follow-up. *Spine* 2006;31:2409-14.
29. Peul WC, van Houwelingen HC, van den Hout WB, et al. Surgery versus prolonged conservative treatment for sciatica. *N Engl J Med* 2007;356:2245-56.
30. Brox JI, Sorensen R, Friis A, et al. Randomized clinical trial of lumbar instrumented fusion and cognitive intervention and exercises in patients with chronic low back pain and disc degeneration. *Spine* 2003;28:1913-21.
31. Brox J, Reikeras O, Nygaard O, et al. Lumbar instrumented fusion compared with cognitive intervention and exercises in patients with chronic back pain after previous surgery for disc herniation: a prospective randomized controlled study. *Pain* 2006;122:145-55.
32. Fairbank J, Frost H, Wilson-MacDonald J, et al. Randomised controlled trial to compare surgical stabilisation of the lumbar spine with an intensive rehabilitation programme for patients with chronic low back pain: the MRC spine stabilisation trial. *BMJ* 2005;330:1233.

33. Fritzell P, Hagg O, Wessberg P, et al. Swedish Lumbar Spine Study Group. 2001 Volvo Award Winner in Clinical Studies: lumbar fusion versus nonsurgical treatment for chronic low back pain: a multicenter randomized controlled trial from the Swedish Lumbar Spine Study Group. *Spine* 2001;26:2521–32.
34. Weber H. Lumbar disc herniation. A controlled, prospective study with ten years of observation. *Spine* 1983;8:131–40.
35. Weinstein JN, Tosteson TD, Lurie JD, et al. Surgical vs. nonoperative treatment for lumbar disk herniation: the Spine Patient Outcomes Research Trial (SPORT): a randomized trial. *JAMA* 2006;296:2441–50.
36. Ekman P, Moller H, Hedlund R. The long-term effect of posterolateral fusion in adult isthmic spondylolisthesis: a randomized controlled study. *Spine J* 2005;5(1):36–44.
37. Moller H, Hedlund R. Surgery versus conservative management in adult isthmic spondylolisthesis—a prospective randomized study: Part 1. *Spine* 2000;25(13):1711–5.
38. Atlas SJ, Deyo RA, Keller RB, et al. The Maine Lumbar Spine Study, Part III. 1-year outcomes of surgical and nonsurgical management of lumbar spinal stenosis. *Spine* 1996;21(15):1787–94.
39. Atlas SJ, Deyo RA, Keller RB, et al. The Maine Lumbar Spine Study, Part II. 1-year outcomes of surgical and nonsurgical management of sciatica. *Spine* 1996;21(15):1777–86.
40. Atlas SJ, Singer DE, Keller RB, et al. Application of outcomes research in occupational low back pain: the Maine Lumbar Spine Study. *Am J Ind Med* 1996;29(6):584–9.
41. Atlas SJ, Keller RB, Robson D, et al. Surgical and nonsurgical management of lumbar spinal stenosis: four-year outcomes from the Maine Lumbar Spine Study. *Spine* 2000;25(5):556–62.
42. Atlas SJ, Keller RB, Wu YA, et al. Long-term outcomes of surgical and nonsurgical management of lumbar spinal stenosis: 8 to 10 year results from the Maine Lumbar Spine Study. *Spine* 2005;30(8):936–43.
43. Chang Y, Singer DE, Wu YA, et al. The effect of surgical and nonsurgical treatment on longitudinal outcomes of lumbar spinal stenosis over 10 years. *J Am Geriatr Soc* 2005;53(5):785–92.
44. Keller RB, Atlas SJ, Singer DE, et al. The Maine Lumbar Spine Study, Part I. Background and concepts. *Spine* 1996;21(15):1769–76.
45. Jhonsson KE, Uden A, Rosen I. The effect of decompression on the natural course of spinal stenosis. A comparison of surgically treated and untreated patients. *Spine* 1991;16(6):615–9.
46. Mariconda M, Fava R, Gatto A, et al. Unilateral laminectomy for bilateral decompression of lumbar spinal stenosis: a prospective comparative study with conservatively treated patients. *J Spinal Disord Tech* 2002;15(1):39–46.
47. Amundsen T, Weber H, Lilleas F, et al. Lumbar spinal stenosis. Clinical and radiologic features. *Spine* 1995;20:1178–86.
48. Birkmeyer NJ, Weinstein JN, Tosteson AN, et al. Design of the Spine Patient Outcomes Research Trial (SPORT). *Spine* 2002;27(12):1361–72.
49. Weinstein JN, Lurie JD, Tosteson TD, et al. Surgical compared with nonoperative treatment for lumbar degenerative spondylolisthesis. Four-year results in the Spine Patient Outcomes Research Trial (SPORT) randomized and observational cohorts. *J Bone Joint Surg Am* 2009;91:1295–304.
50. Zucherman JF, Hsu KY, Hartjen CA, et al. A prospective randomized multicenter study for the treatment of lumbar spinal stenosis with the X STOP interspinous implant: 1-year results. *Eur Spine J* 2004;13(1):22–31.
51. Hsu KY, Zucherman JF, Hartjen CA, et al. Quality of life of lumbar stenosis-treated patients in whom the X STOP interspinous device was implanted. *J Neurosurg Spine* 2006;5(6):500–7.
52. Anderson PA, Tribus CB, Kitchel SH. Treatment of neurogenic claudication by interspinous decompression: application of the X STOP device in patients with lumbar degenerative spondylolisthesis. *J Neurosurg Spine* 2006;4(6):463–71.
53. Van der Roer N, Ostelo RWJG, Bekkering G, et al. Minimal clinically important change for pain intensity, functional status, and general health status in patients with nonspecific low back pain. *Spine* 2006;31:578–82.
54. Kovacs F, Abraira V, Royuela A, et al. Minimal clinically important change for pain intensity and disability in patients with nonspecific low back pain. *Spine* 2007;32:2915–20.
55. Deyo RA, Mirza SK, Martin BI, et al. Trends, major medical complications and charges associated with surgery for lumbar spinal stenosis in older adults. *JAMA* 2010;303(13):1259–65.
56. Park DK, An HS, Lurie JD, et al. Does multilevel lumbar stenosis lead to poorer outcomes? A subanalysis of the Spine Patient Outcomes Research Trial (SPORT) lumbar stenosis study. *Spine* 2010;35:439–46.
57. Jansson KA, Németh G, Granath F, et al. Health-related quality of life (EQ-5D) before and one year after surgery for lumbar spinal stenosis. *J Bone Joint Surg Br* 2009;91(2):210–6.
58. Fischgrund JS, Mackay M, Herkowitz HN, et al. Degenerative lumbar spondylolisthesis with spinal stenosis: a prospective, randomized study comparing decompressive laminectomy and arthrodesis with and without spinal instrumentation. *Spine* 1997;22(24):2807–12.
59. Herkowitz HN, Kurz LT. Degenerative lumbar spondylolisthesis with spinal stenosis: a prospective study comparing decompression with decompression and intertransverse process arthrodesis. *J Bone Joint Surg Am* 1991;73(6):802–8.
60. Grob D, Humke T, Dvorak J. Degenerative lumbar spinal stenosis: decompression with and without arthrodesis. *J Bone Joint Surg Am* 1995;77(7):1036–41.
61. Abdu WA, Lurie JD, Spratt KF, et al. Degenerative spondylolisthesis. Does fusion method influence outcome? Four-year results of the Spine Patient Outcomes Research Trial. *Spine* 2009;34:2351–60.
62. Thomsen K, Christensen FB, Eiskjaer SP, et al. The effect of pedicle screw instrumentation on functional outcome and fusion rates in posterolateral lumbar spinal fusion: a prospective randomized clinical study. *Spine* 1997;22(24):2813–22.
63. Fritzell P, Hagg O, Nordwall A. Swedish Lumbar Spine Study Group. Complications in lumbar fusion surgery for chronic low back pain: comparison of three surgical techniques used in a prospective randomized study. *Eur Spine J* 2003;12(2):178–89.
64. Mirza SK, Deyo RA, Heagerty PJ, et al. Development of an index to characterize the “invasiveness” of spine surgery: validation by comparison to blood loss and operative time. *Spine* 2008;33(24):2651–61.
65. Moojen WA, Arts MP, Brand R, et al. The Felix-trial. Double-blind randomization of interspinous implant or bony decompression for treatment of spinal stenosis related intermittent neurogenic claudication. *BMC Musculoskelet Disord* 2010;27(11):100.
66. Carragee EJ, Deyo RA, Kovacs FM, et al. Clinical research. Is the spine field a mine field? *Spine* 2009;34:423–30.
67. Kohles SS, Kohles DA, Karp AP, et al. Time-dependent surgical outcomes following cauda equina syndrome diagnosis. Comments on a meta-analysis. *Spine* 2004;29:1281–7.



Review Article

The efficacy, safety, effectiveness, and cost-effectiveness of ultrasound and shock wave therapies for low back pain: a systematic review

Jesús Seco, MD, PhD^{a,b}, Francisco M. Kovacs, MD, PhD^{b,c,*}, Gerard Urrutia, MD^{b,d,e}

^aDepartment of Nursing and Physical Therapy, Institute of Biomedicine (IBIOMED), University of León, Ponferrada, Spain

^bSpanish Back Pain Research Network, Paseo Mallorca 36, 07012 Palma de Mallorca, Spain

^cScientific Department, Fundació Kovacs, Palma de Mallorca, Paseo Mallorca 36, 07012 Palma de Mallorca, Spain

^dIberoamerican Cochrane Center-Servei d'Epidemiologia Clínica i Salut Pública, Institut d'Investigació Biomèdica Sant Pau, Hospital de Sant Pau, C/ Sant Antoni Maria Claret, 167, 08025 Barcelona, Spain

^eCIBER Epidemiología y Salud Pública (CIBERESP), Spain

Received 3 September 2010; revised 12 December 2010; accepted 7 February 2011

Abstract

BACKGROUND CONTEXT: Shock wave and especially ultrasound are commonly used to treat low back pain (LBP) in routine practice.

PURPOSE: To assess the evidence on the efficacy, effectiveness, cost-effectiveness, and safety of ultrasound and shock wave to treat LBP.

STUDY DESIGN: Systematic review.

METHODS: An electronic search was performed in MEDLINE, EMBASE, and the Cochrane Library databases up to July 2009 to identify randomized controlled trials (RCTs) comparing vibrotherapy with placebo or with other treatments for LBP. No language restrictions were applied. Additional data were requested from the authors of the original studies. The risk of bias of each study was assessed following the criteria recommended by the Cochrane Back Review Group.

RESULTS: Thirteen studies were identified. The four RCTs complying with the inclusion criteria included 252 patients. Two of the three RCTs on ultrasound had a high risk of bias. For acute patients with LBP and leg pain attributed to disk herniation, ultrasound, traction, and low-power laser obtained similar results. For chronic LBP patients without leg pain, ultrasound was less effective than spinal manipulation, whereas a shock wave device and transcutaneous electrical nerve stimulation led to similar results. Results from the only study comparing ultrasound versus a sham procedure are unreliable because of the inappropriateness of the sham procedure, low sample size, and lack of adjustment for potential confounders. No study assessed cost-effectiveness. No adverse events were reported.

CONCLUSION: The available evidence does not support the effectiveness of ultrasound or shock wave for treating LBP. High-quality RCTs are needed to assess their efficacy versus appropriate sham procedures, and their effectiveness and cost-effectiveness versus other procedures shown to be effective for LBP. In the absence of such evidence, the clinical use of these forms of treatment is not justified and should be discouraged. © 2011 Elsevier Inc. All rights reserved.

Keywords:

Low back pain; Ultrasound; Vibrotherapy; Shock wave; Systematic review; Effectiveness

Introduction

Nonspecific or common low back pain (LBP) is defined as pain between the costal margins and the inferior gluteal folds, which may be associated with pain referred down to the leg ("leg pain"), and is usually accompanied by painful

limitation of movement [1]. Diagnosing common LBP implies that the pain is not related to conditions such as fractures, spondylitis, direct trauma, or neoplastic, infectious, vascular, metabolic, or endocrine-related processes [1].

Two forms of mechanical vibration are used for treating LBP. Ultrasound is the term used when the frequency of

FDA device/drug status: not applicable.

Author disclosures: none.

Gerard Urrutia is a PhD candidate at the Public Health and Research Methodology Programme, Universitat Autònoma de Barcelona (UAB).

* Corresponding author. Departamento Científico, Fundació Kovacs, Paseo Mallorca 36, 07012 Palma de Mallorca, Spain. Tel.: (34) 971-720809; fax: (34) 971-720774.

E-mail address: kovacs@kovacs.org (F.M. Kovacs)

vibration is above 20,000 Hz and generates heat. Ultrasound penetrates the tissue in a focused directional manner, transmitting heat deep into the tissue. It is believed that ultrasound can increase local metabolism and blood circulation, enhance the flexibility of connective tissue, and accelerate tissue regeneration, potentially reducing pain and stiffness, while improving mobility [2,3]. Shock wave is a form of treatment that applies vibration at a low frequency (10, 50, 100, or 250 Hz), causing an oscillatory pressure. Several devices have been designed to provide “shock waves,” such as “FairMed,” “muscle relaxation machines,” and vibratory platforms. Shock wave is commonly referred to as “vibrotherapy,” but it will be referred to as shock wave hereafter because both shock wave and ultrasound use vibration with a therapeutic goal and, therefore, strictly speaking, both could be referred to as “vibrotherapy.”

The most recent evidence-based guidelines for the treatment of LBP do not recommend shock wave or ultrasound [4–6]. Nevertheless, ultrasound is commonly used in routine clinical practice for musculoskeletal problems, including LBP [7–9]. Approximately 50% of UK physiotherapists, 65% of US physiotherapists, and 94% of Canadian physiotherapists use it [10]. In the United States, 55% of primary care practitioners recommend ultrasound as a form of treatment [11]. Moreover, the most recent Cochrane review on ultrasound suggests that it might be useful for treating knee osteoarthritis, although firm conclusions could not be drawn because of the low quality of the trials included [12].

No systematic reviews on shock wave have been published, and the last one on ultrasound for LBP did not find any randomized clinical trials. However, it dates back to 2001 [13], and new evidence may have been published since then.

Therefore, the objective of this study was to systematically review the evidence on the efficacy, safety, effectiveness, and cost-effectiveness of “ultrasound/shock wave” versus placebo or other treatments, for treating pain and disability in LBP patients with or without radiculopathy.

Materials and methods

The protocol of this study was approved by and registered in the Spanish Ministry of Health’s (“Fondo de Investigación Sanitaria” FIS 03/0908) and the Spanish Back Pain Research Network’s databases.

Search and study selection

An electronic search of studies was performed in CENTRAL (The Cochrane Library 2009; Issue 4), MEDLINE (1966–2008), and EMBASE (1980–2008). The search strategy was designed to ensure maximum sensitivity and did not apply language restrictions. It is shown in the [Appendix](#).

Studies were included in this review if they were randomized controlled trials (RCTs); compared ultrasound or shock wave with any kind of active or inert procedure; included patients with LBP with or without radicular or referred pain down to the leg (“leg pain”); and assessed at least one of the following variables: pain severity, disability, quality of life, daily activities, impact on work, patient satisfaction, medication use, side effects, health resources employed, and total costs. No language restrictions were applied.

References identified through the electronic search were screened. The full text of those that were eligible was assessed for inclusion criteria independently by two authors (JS and GU). Disagreements were resolved by consensus with a third author (FMK). References of potential additional studies were searched for in the reference section of the studies included.

Risk of bias assessment and data analysis

When necessary, unpublished data were requested from the authors of the original studies. Details are shown in [Tables 1–3](#).

Following the recommendations of the Cochrane Back Review Group, the risk of bias of the studies included was independently assessed by two of the reviewers (JS and GU), and it was agreed that any disagreements would be solved by consensus with the third author [14]. According to these criteria, a study was rated as having a “low risk of bias” when it met at least 6 of the 12 proposed criteria. For studies that compared procedures that did not allow masking of therapists and patients, these two criteria were scored as “not applicable.” For the latter, the maximum possible score was 10, and they were considered as having a “low risk of bias” if they scored ≥ 5 . However, studies with serious flaws were considered as having a “high risk of bias” independently of their score [14].

Data extraction was undertaken independently and in duplicate, using standardized forms. For each study, data on pain severity, disability, quality of life, daily activities, impact on work, patient satisfaction, medication use, side effects, health resources employed, and total costs, were searched for, and all data on all the variables gathered in each individual study were extracted.

A quantitative synthesis of data (meta-analysis) was not possible because of the variability of outcome measures and heterogeneity of the methods used across the studies. Therefore, a qualitative analysis of data was performed, based on the methodological quality of the trials included and consistency of their findings.

Role of the funding sources

The funding institutions had no role in the design and conduction of the study; data collection, management, analysis, and interpretation; preparation, review, and approval

Table 1
Characteristics of the studies that were included

Study	Methods	Participants	Interventions	Outcomes
Ansari, 2006 (Iran) [18]	RCT. An orthopedist who was unaware of the treatment schedule selected patients who were referred to a single physiotherapist. She allocated patients using sealed opaque envelopes withdrawn by chance. Single blind (patients).	Fifteen patients with chronic nonspecific low back pain without radiating pain. Five dropped out because of noncompliance and discontinuation of treatment, and only 10 were analyzed. Patients were not comparable at baseline in age (ultrasound: 35.5 ± 13.8 y; placebo: 26.4 ± 11.3 y), gender (ultrasound: all men; placebo: two men and three women), and body mass index (ultrasound: 28.3 ± 2.9 kg/m ² ; placebo: 23.8 ± 5.2 kg/m ²) ($p = .032$).	All treatments were applied over a period of 3 wk, 3 days a week (10 sessions in total). Therapeutic ultrasound (n=5): using 1 MHz with a 5-cm ² sound head at an intensity of 1.5 W/cm ² , in continuous mode, on the right and left sides of the posterior lumbar region, for 8 min. Placebo group (n=5): The same ultrasound device was used, but it was switched off. The application pattern (in terms of duration and location) was the same as in the active group. Co-interventions: Patients were advised to continue previous treatment but not to initiate any new treatments or take new analgesics. No exercise programs were prescribed.	Follow up: 3 wk. Functional Rating Index (0–100): 10 items (pain, sleeping, personal care, travel, work, activity, frequency of pain, lifting, walking, and standing). Lumbar range of motion: extension, flexion, and lateral flexion. Electroneurophysiological evaluation: <ul style="list-style-type: none"> • H reflex latency and H_{max}/M_{max}. • Impedance of less than 10 kΩ • Rectangular pulses (p), duration 0.5 ms, rate 1p/3s • In the soleus muscle • In both sides.
Unlu, 2008 (Turkey) [17]	RCT. All patients were sequentially recruited by the same physiatrist, who also performed clinical assessments during follow-up visits. No details about randomization. Single blind (outcome assessor).	60 patients (18 men and 42 women) with a mean age of 44.5 y (range, 20–60 y) with acute low back and sciatica or femoral neuralgia symptoms for less than 3 mo because of the herniation of one or more lumbar disks, documented by magnetic resonance imaging. No dropouts. No significant baseline differences between groups.	All treatments were applied over a period of 3 wk; once a day, 5 days a week (15 sessions in total). Lumbar traction (n=20): standard motorized traction therapy system (15 min per session). Therapeutic ultrasound (n=20): using 1 MHz with a 5-cm ² sound head, at an intensity of 1.5 W/cm ² , in continuous mode, on the right and left sides of the posterior lumbar region, for 8 min. Low-power laser (n=20): a Gal-Al-As diode, at a power output of 50 mV and wavelength of 830 nm. The dose at each point was 1 J. Time: 4 min at each point. Co-interventions: “Other co-interventions were not allowed during the treatment period.”	Follow-up: 4 and 12 wk after treatment. Lumbar range of motion: lateral flexion and Schober test. Muscular tenderness on palpation of paravertebral muscles. Straight leg raising test ($\leq 70^\circ$ in the affected leg). Pain (low back pain and radiating leg pain) (VAS 0–100). Roland Disability Questionnaire (0–24). Modified Oswestry Disability Index (0–50).
Mohseni-Bandpei, 2006 (UK) [16]	RCT. Patients were assigned a number according to a block-style randomization scheme.	120 patients aged between 18 and 55 y with pain between L1 and L5 and the sacroiliac joints for ≥ 3 mo,	Treatment sessions lasted for ≈ 20 min.	Follow-up: End of treatment (3–4 wk), in person, and 6 mo later, through a postal questionnaire.

(continued)

Table 1 (continued)

Study	Methods	Participants	Interventions	Outcomes
	Single blind (outcome assessor).	without sciatic pain, nerve root compression, or disk herniation. Mean (SD) age (assessment after treatment completion/6 mo later): 34.8 (10.2) y/37.2 (10.6) y. Men, n (%): 22 (39%)/24 (43%). 112 patients were assessed after treatment completion and 73 patients were assessed 6 mo later. No significant baseline differences between groups.	Spinal manipulation (n=56): a dynamic, short-lever, high-velocity thrust was applied exerting a force on the lumbar spine and sacroiliac joint. Four sessions (range, two to seven sessions), once or twice per week. Ultrasound (n=56): a frequency of 1 MHz; continuous pattern ultrasound was applied to the lower back, on the painful area; an intensity of between 1.5 and 2.5 W/cm ² ; 5–10 min depending on the size of the treatment area; six sessions (range, 3–11), once or twice per week. Co-interventions: A written set of exercises were generated by the PhysioTools computer package, and the physiotherapist selected some of them for each patient. Medications (eg, NSAID, pain killers, muscle relaxants) were allowed and registered in both groups.	Pain intensity during the last week (VAS 0–10) [42]. Disability (Oswestry Disability Index) [43]. Range of motion (modified-modified Schober test) [44]. Surface electromyography of multifidus and iliocostalis: Median frequency, and median frequency slope (rate of shift calculated from the median frequency calculated for every second of recording, which was considered to be an index of muscle fatigue). Muscle endurance (Biering-Sorensen test) [45].
Barker, 2008 (United Kingdom) [15]	RCT. Noninferiority trial. Randomization was performed using a random numbers table with permuted blocks. Consecutively numbered, sealed, opaque envelopes were used. Single blind (assessors).	Among 120 eligible adult patients older than 18 y, with low back pain for ≥3 mo and no leg pain, 60 were referred to a single physiotherapy department. Mean (SD) age: 53.4 (±11.5) y, 50% men. Six (10%) patients withdrew (one in the TENS group and five in the FairMed). No significant baseline differences between groups.	All treatments were applied over a period of 3 wk. TENS (n=27): Continuous trains at high frequency (80 Hz, using square-wave 100-μs pulses). Two surface electrodes (5 cm×5 cm ² TPN 40 each) were placed in or adjacent to the painful area at a distance of 5 to 20 cm apart. The intensity was adjusted to produce a tingling sensation that was approximately two to three times the sensory threshold. FairMed device (n=27): Subjects were asked to use the device for 30-min sessions. Co-interventions: No co-interventions were allowed.*	Follow-up: 3 wk. Pain (VAS) [46]. Disability (Oswestry Disability Index v. 2.1) [47], functional physical tests [48] (5-min walking distance, 1-min stair climb, and 1-min standing up from and sitting down on a chair). Emotional functioning (Hospital Anxiety and Depression Scale) [49]. Fear of movement (Tampa Scale for Kinesiophobia) [50]. Catastrophizing (Pain Coping Scale) [51]. Self-efficacy beliefs (Pain Self-Efficacy Questionnaire) [52]. Global rating of patient reported improvement and satisfaction (Patient Global Impression of Change scale) [53]. Adverse events.

VAS, visual analog scale; SD, standard deviation; NSAID, nonsteroidal anti-inflammatory drug; TENS, transcutaneous electrical nerve stimulation.

* Personal communication from Ms Barker.

Table 2
Risk of bias of the studies included [14]

Sources of risk of bias	Ansari (2006) [18]	Unlu (2008) [17]	Mohseni-Bandpei (2006) [16]	Barker (2008) [15]
A 1. Was the method of randomization adequate?	Yes*	Unsure (not described)	Yes [†]	Yes
B 2. Was treatment allocation concealed?	Yes*	Unsure	Yes [†]	Yes
C Was knowledge of the allocated interventions, adequately prevented during the study?				
3. Was the patient blinded to the intervention?	Unsure [‡]	Not applicable	Not applicable	Not applicable
4. Was the care provider blinded to the intervention?	Not applicable	Not applicable	Not applicable	Not applicable
5. Was the outcome assessor blinded to the intervention?	No	Yes	Yes	Yes
D Were incomplete outcome data adequately addressed?				
6. Was the dropout rate described and acceptable?	No	Yes [§]	Yes (immediately after the end of treatment) No (at 6 mo)	Yes [¶]
7. Were all randomized participants analyzed in the group to which they were allocated?	No	Yes	Yes (immediately after the end of treatment) No (at 6 mo)	Yes
E 8. Are reports of the study free of suggestion of selective outcome reporting?	Yes	Yes	Yes	Yes
F Other sources of potential bias:				
9. Were the groups similar at baseline regarding the most important prognostic indicators?	No	Yes	Yes	Yes
10. Were co-interventions avoided or comparable?	Yes*	Yes	Yes [†]	Yes [#]
11. Was the compliance acceptable in all groups?	Yes	Yes	Yes	Yes
12. Was the timing of the outcome assessment in all groups similar?	Yes	Yes	Yes	Yes
Total	6/11**	8/10	10/10 (for results from assessment immediately after treatment) 8/10 (for results at 6 mo)**	10/10

TENS, transcutaneous electrical nerve stimulation.

* Details not available in the article, which were provided by Dr Nouredin Ansari on request from the authors of this review.

† Details not available in the article, which were provided by Dr Mohseni-Bandpei on request from the authors of this review.

‡ No co-interventions were allowed (data provided by Ms Barker and Dr Fairbank on request from the authors of this review).

§ Although patients were supposed to be masked, sham intervention did not cause heat (whereas active ultrasound does). This might have compromised patients' blindness.

|| No patients dropped out; all patients were analyzed.

¶ The sample was established anticipating for $\leq 20\%$ losses to follow-up. Immediately after treatment, eight of 112 patients (6.6%) were lost to follow-up. At the 6-month assessment (through a postal questionnaire), 47 patients (42%) were lost. Although a sensitivity analysis assuming different values for these patients was not performed, there was no crossover, and all patients were analyzed in the group they had been assigned to.

One patient in the TENS group (of 28), and five (of 32) in the FairMed dropped out. Although a sensitivity analysis assuming different values for these patients was not performed, there was no crossover, and all patients were analyzed in the group they had been assigned to.

** Despite this score, this study had serious flaws implying a high risk of bias.

of the manuscript; or in the decision to submit the article for publication.

Results

The electronic search provided 1,419 references, which led to the identification of 13 studies potentially eligible for inclusion in this review [15–27]. Twelve focused on ultrasound [16–27], and one on a shock wave device [15]. Nine studies were excluded: two because they turned out to not be randomized studies [20,21], one

because it used sham ultrasound (a dose close to 0–0.1 W/cm²—intermittently and without generating a heat effect) as a “placebo” treatment to be compared with usual treatment and exercise [19], and six because they used ultrasound in combination with other treatments in all the groups, making it impossible to isolate its effect [22–27].

The remaining four studies were included in this systematic review [15–18]. They included a total of 242 patients, and their characteristics are shown in Table 1. There was no disagreement between reviewers with regard to data extraction and the assessment of risk of bias.

Table 3
Summary of results

Study	Pain	Disability	Mobility	Other
Ansari 2006 [18]	Not assessed	FRI (%): mean (SD) <ul style="list-style-type: none"> • Ultrasound: before treatment, 56.5 (20.35); after treatment, 34.5 (13.5) • Placebo: before treatment, 46.95 (14.38); after treatment, 39.9 (16.5) 	Range of motion (degree): mean (SD) Flexion <ul style="list-style-type: none"> • Ultrasound: before treatment, 117.4 (2.5); after treatment, 128.6 (14.3) • Placebo: before treatment, 103.4 (13.39); after treatment, 109.2 (10.6) Extension <ul style="list-style-type: none"> • Ultrasound: before treatment, 23.8 (4.15); after treatment, 30 (6.4) • Placebo: before treatment, 27.2 (3.03); after treatment, 29 (4.2) Right lateral flexion <ul style="list-style-type: none"> • Ultrasound: before treatment, 20.2 (7.3); after treatment, 28 (8.6) • Placebo: before treatment, 22.6 (5.2); after treatment, 26 (4.8) Left lateral flexion <ul style="list-style-type: none"> • Ultrasound: before treatment, 22 (7.77); after treatment, 28 (7.1) • Placebo: before treatment, 23.8 (6.53); after treatment, 28.4 (4) 	Electromyography features: mean (SD) Right H_{max}/M_{max} <ul style="list-style-type: none"> • Ultrasound: before treatment, 0.36 (0.07); after treatment, 0.36 (0.08) • Placebo: before treatment, 0.38 (0.04); after treatment, 0.42 (0.01) Left H_{max}/M_{max} <ul style="list-style-type: none"> • Ultrasound: before treatment, 0.34 (0.1); after treatment, 0.35 (0.09) • Placebo: before treatment, 0.41 (0.05); after treatment, 0.44 (0.03) Right latency (ms) <ul style="list-style-type: none"> • Ultrasound: before treatment, 27.32 (1.1); after treatment, 27.7 (1.1) • Placebo: before treatment, 26.4 (2.7); after treatment, 26.5 (2.7) Left latency (ms) <ul style="list-style-type: none"> • Ultrasound: before treatment, 27.7 (0.9); after treatment, 28.1 (1.7) • Placebo: before treatment, 26.11 (2.8); after treatment, 26.5 (2.5)
Unlu 2008 [17]	LBP (VAS 0–100): mean (SD): <ul style="list-style-type: none"> • Traction: before treatment, 58.2 (18.1); at 3 mo, 31.3 (16.4) • Laser: before treatment, 54.0 (17.0); at 3 mo, 30.0 (16.8) • Ultrasound: before treatment, 51.7 (18.7); at 3 mo, 26.9 (15.2) Radicular pain (VAS 0–100): mean (SD) <ul style="list-style-type: none"> • Traction: before treatment, 59.6 (15.4); at 3 mo, 29.5 (16.7) • Laser: before treatment, 53.1 (25.9); at 3 mo, 23.6 (17.7) • Ultrasound: before treatment, 56.0 (15.3); at 3 mo, 25.2 (13.9) 	RDQ: mean (SD) <ul style="list-style-type: none"> • Traction: before treatment, 14.2 (4.3); at 3 mo, 8.9 (4.0) • Laser: before treatment, 12.5 (5.0); at 3 mo, 6.7 (4.5) • Ultrasound: before treatment, 13.4 (4.5); at 3 mo, 8.6 (6.0) MODQ: mean (SD) <ul style="list-style-type: none"> • Traction: before treatment, 19.3 (5.3); at 3 mo, 14.9 (4.9) • Laser: before treatment, 18.4 (7.1); at 3 mo, 13.6 (6.2) • Ultrasound: before treatment, 19.6 (6.4); at 3 mo, 14.4 (5.9) 	Lateral flexion: mean (SD) <ul style="list-style-type: none"> • Traction: before treatment, 47.0 (3.3); at 3 mo, 46.1 (4.2) • Laser: before treatment, 47.0 (5.1); at 3 mo, 45.2 (4.7) • Ultrasound: before treatment, 49.1 (4.5); at 3 mo, 47.8 (3.2) Schober test: mean (SD) <ul style="list-style-type: none"> • Traction: before treatment, 4.0 (0.9); at 3 mo, 4.2 (0.8) • Laser: before treatment, 4.1 (0.9); at 3 mo, 4.3 (0.9) • Ultrasound: before treatment, 4.2 (1.0); at 3 mo, 4.5 (0.9) 	Muscular tenderness (0–4): mean (SD) <ul style="list-style-type: none"> • Traction: before treatment, 0.7 (0.9); at 3 mo, 0.3 (0.5) • Laser: before treatment, 0.8 (0.8); at 3 mo, 0.5 (0.6) • Ultrasound: before treatment, 1.2 (1.1); at 3 mo, 0.5 (0.7) Straight leg raising test (degree): median (SD) <ul style="list-style-type: none"> • Traction: before treatment, 55.2 (7.1); at 3 mo, 57.5 (13.8) • Laser: before treatment, 55.2 (9.6); at 3 mo, 65.0 (5.7) • Ultrasound: before treatment, 54.4 (8.9); at 3 mo 57.0 (10.5)
Mohseni-Bandpei 2006 [16]	VAS <ul style="list-style-type: none"> • Baseline data: mean (SD): Manipulation: 65 (19) Ultrasound: 63 (19) 	ODI (%) <ul style="list-style-type: none"> • Baseline data: mean (SD): Manipulation: 30.8 (12.7) 	Lumbar flexion (mm) <ul style="list-style-type: none"> • Baseline data: mean (SD): Manipulation: 47 (13) Ultrasound: 51 (11) 	Median frequency multifidus (Hz) (mean): <ul style="list-style-type: none"> • Baseline data: mean (SD): Manipulation: 105 (16)

	<ul style="list-style-type: none"> Improvement: mean (95% CI): Difference in favor of manipulation at 6 mo follow-up: 15.1 (7.6, 22.6) 	<p>Ultrasound: before treatment, 32.2 (14.9)</p> <ul style="list-style-type: none"> Improvement: mean (95% CI): Difference in favor of manipulation at 6 mo follow-up: 5.2 (2.6, 7.8) 	<ul style="list-style-type: none"> Improvement mean (95% CI): Differences in favor of manipulation between the end of treatment and the 6 mo follow-up: 9.4 (5.5, 13.4) <p>Lumbar extension (mm)</p> <ul style="list-style-type: none"> Baseline data: mean (SD): Manipulation: 11 (7) Ultrasound: 10 (7) Improvement mean (95% CI): Differences in favor of manipulation between the end of treatment and the 6 mo follow-up: 3.4 (1.0, 5.8) 	<p>Ultrasound: 99 (16)</p> <ul style="list-style-type: none"> Improvement mean (95% CI): Differences in favor of manipulation between the end of treatment and the 6 mo follow-up: 6.8 (−1.2, 14.9) <p>Median frequency iliocostalis (Hz):</p> <ul style="list-style-type: none"> Baseline data: mean (SD): Manipulation: 67 (7), Ultrasound: 72 (7) Improvement mean (95% CI): Differences in favor of manipulation between the end of treatment and the 6 mo follow-up: 2.4 (−2.5, 7.1) <p>Median frequency slope multifidus</p> <ul style="list-style-type: none"> Baseline data: mean (SD): Manipulation: −0.65 (0.18), Ultrasound: −0.54 (0.25) Improvement mean (95% CI): Differences in favor of manipulation between the end of treatment and the 6 mo follow-up: 0.3 (0.1, 0.5) <p>Median frequency slope multifidus</p> <ul style="list-style-type: none"> Baseline data: mean (SD): Manipulation: −0.35 (0.27), Ultrasound: −0.34 (0.37) Improvement mean (95% CI): Differences in favor of manipulation between the end of treatment and the 6 mo follow-up: 0.1 (−0.1, 0.5) <p>Co-interventions* (% of patients using each kind of drug)</p> <ul style="list-style-type: none"> Manipulation: Pain killers (14%), NSAID (34%), muscle relaxant (0%), none (52%) Ultrasound: Pain killers (23%), NSAID (13%), muscle relaxant (2%), none (62%) <p>Adverse events: None detected*</p> <p>TSK; mean (SD)</p> <ul style="list-style-type: none"> FairMed: −1.8 (5.8) TENS: −2.0 (7.7) <p>HAD-A; mean (SD)</p> <ul style="list-style-type: none"> FairMed: −1.4 (2.7) TENS: −0.3 (3.0) <p>HAD-D; mean (SD)</p> <ul style="list-style-type: none"> FairMed: −0.2 (2.0)
Barker 2008 [15]	<p>(Change in outcome measures between baseline and 3 wk)</p> <p>VAS; mean (SD)</p> <ul style="list-style-type: none"> FairMed: −8 (18) TENS: −7 (14) <p>PSE; mean (SD)</p> <ul style="list-style-type: none"> FairMed: 1.9 (7.8) TENS: 4.4 (7.5) 	<p>(Change in outcome measures between baseline and 3 wk)</p> <p>ODI (0–100); mean (SD)</p> <ul style="list-style-type: none"> FairMed: −0.6 (8.7) TENS: −0.9 (5.1) 	<p>(Change in outcome measures between baseline and 3 wk)</p> <p>Sit to stand; mean (SD)</p> <ul style="list-style-type: none"> FairMed: 1.0 (2.1) TENS: 1.0 (2.2) <p>Stairs; mean (SD)</p> <ul style="list-style-type: none"> FairMed: 0.4 (1.9) TENS: 0.5 (1.0) 	

(continued)

Table 3 (continued)

Study	Pain	Disability	Mobility	Other
	PCS; mean (SD) • FairMed: -1.6 (5.4) • TENS: 2.0 (7.7)		Walking distance (meters in 5 min); mean (SD) • FairMed: 3.1 (41.4) • TENS: 9.1 (-6.4)	• TENS: -0.6 (1.9) PGIC (%) • FairMed: Improved: 27%, unchanged: 73% • TENS: Improved: 45%, unchanged: 44%, worsened: 11% Adverse events: None detected [†]

Electromyography parameters, H reflex (range, 25–32 ms); FRI, Functional Rating Index (range from better to worse; 0–100%); HADS-A, Hospital Anxiety and Depression Scale—Anxiety (range, 0–21); HADS-D, Hospital Anxiety and Depression Scale—Depression (range, 0–21); Median frequency multifidus (Hz); range 20 to 450 Hz; MODI, Modified Oswestry Disability Index (range from better to worse; 0–100%); ODI, Oswestry Disability Index (range from better to worse; 0–100%); PGIC, Patient Global Impression of Change scale; patients scored their ability to cope with pain, and to perform everyday activities, but results were only given categorized as “improved,” “unchanged,” or “worsened”; PCS, Pain Catastrophizing Scale (range, 0–52); PSE, Pain Self-Efficacy (range, 0–60); RDQ, Roland Disability Questionnaire (range from better to worse; 0–24); ROM, range of motion (range, degrees); Flexion (range, 0°–75°); Extension (range, 0°–30°); Lateral flexion (range, 0°–35°); Schober test (range, from -9 to +15 cm); Sit to stand (N°/min); SLRT, straight leg raising test (range, 0 to >70°); Stairs (N°/min); TSK, Tampa Scale for kinesiophobia (range, 17–60); VAS, visual analog scale (range from better to worse; 0–100); Walking distance (meters in 5 min); SD, standard deviation; 95% CI, 95% confidence interval; NSAID, nonsteroidal anti-inflammatory drug; TENS, transcutaneous electrical nerve stimulation.

Data in bold italics indicate the results that are statistically significant.

* Personal communication from Dr Mahsemi-Bandpei (data in the original article were misprinted or absent).

† Personal communication from Dr Fairbank and Ms Barker (data absent in the original article).

Across the studies, ultrasound was compared with a sham procedure (a deactivated device that was moved across the painful area, but which did not produce heat) [18], lumbar traction [17], low-power laser [17], and spinal manipulation [16]. Shock wave therapy was applied by using the “FairMed” device [15] and was compared with transcutaneous electrical nerve stimulation (TENS). Each of these interventions is described in detail in Table 1.

The three RCTs on ultrasound included a total of 182 patients, of whom 71 (39.0%) were men, with a mean (standard deviation) age of 37.5 (11.47) years [16–18]. The RCT on shock wave included 60 patients, of whom 30 (50%) were men, with a mean (standard deviation) age of 53.4 (11.5) years [15] (Table 1). In general, patients in these studies had moderate degrees of pain and disability.

All studies measured disability as an outcome measure, although instruments used to assess it varied across studies (Table 1). Other outcome measures were pain, mobility (flexion, extension, lateral flexion, and Schober test), physical functioning (sit to stand, climbing stairs, walking distance), electrophysiological parameters, pain self-efficacy, pain coping, kinesiophobia, anxiety, and depression scale (Table 1). Follow-up ranged between 3 weeks and 6 months, and losses to follow-up were 0% [17], 10% [15], 25.6% [16], and 33% [18] (Table 1).

The assessment of risk of bias is shown in Table 2. If such an assessment had been based only on the information available in the original report, the scores of two of the studies would have been significantly lower [16,18], and one of them would have been classified as having a high risk of bias [18]. Taking into account the additional information provided by the authors of the original studies, all of them scored as having a low risk of bias [15–18]. However, despite these scores, two of the studies had serious flaws and were considered as having a high risk of bias (Table 2) [16,18].

All studies were single blinded. Because of the nature of the interventions being compared, blinding of patients and therapists was considered to be feasible only in one study, which failed to do so [18]. Only the article on shock wave described the procedure for randomization and allocation concealment [15], but the authors of other two provided these data on request [16,18]. The study on shock wave was the only one for which one of the authors reported a potential conflict of interest [15].

In 60 acute patients with LBP and leg pain attributed to lumbar disk herniations, lumbar traction, ultrasound, and low-power laser led to the same results on pain and disability at 1 and 3 months, albeit the improvement in the straight leg raising test was approximately 7° greater in the laser group than in the ultrasound one (Table 3) [17].

The other three studies focused on chronic (≥3 months) non-specific LBP without leg pain. In a study with 15 patients, in which five were lost to follow-up, at 3 weeks disability had improved approximately 10% more in patients undergoing ultrasound than in those in whom a deactivated

device was applied. This was the study with the lowest quality among those included in this review. It did not assess pain severity, and there were no differences in flexion, extension, lateral flexion, and the H reflex [18].

In a study with 112 patients, at 6 months, manipulation led to significantly greater improvements than ultrasound in: pain (an additional benefit of approximately 1.5 visual analog scale points), disability (approximately 5 Oswestry Disability Index points), flexion (modified-modified Schober test approximately 9 mm), extension (approximately 3 mm), and median frequency slope in the multifidus, whereas there were no differences between groups in the median frequency slope of surface electromyography signals for iliocostalis lumborum or in the median frequency for multifidus and iliocostalis muscles (Table 3) [16].

In a study with 60 patients, at 3 weeks, shock wave led to results that were similar to those from TENS with regard to pain, disability, kinesiophobia, pain self-efficacy, pain coping scale, health anxiety and depression scale, and physical functioning. In this study, the number of patients reporting variations in their perception of ability to cope with pain was greater in the TENS group than in the FairMed group: five (18%) more patients reported improvement, and three (11%) more reported worsening (Table 3) [15].

No data on costs or cost-effectiveness were gathered in these studies. No data on side effects or adverse events were reported but, on request, the authors of three of the studies reported that there were none [15,16,18].

Discussion

Results from this review do not support the use of ultrasound or shock wave for treating patients with LBP and leg pain. Only one study included such patients in which leg pain was attributed to disk herniation and ultrasound was compared with low-power laser and traction [17]. Neither traction nor low-power laser has shown to be effective for LBP, with or without sciatica [28,29]. Therefore, results suggesting that ultrasound is equivalent to these procedures for patients with lumbar disk herniation cannot be seen as supporting the effectiveness of any of those techniques [17]. In fact, the clinical evolution of patients in that particular study may simply reflect the natural history of acute disk herniation (Table 3) [17].

Results from this review do not support the clinical use of ultrasound for patients with common LBP without leg pain, either. One study assessed the effect of ultrasound versus a sham procedure [18]. However, neither therapists nor outcome assessors were masked, and the sham procedure did not produce heat, which may have compromised patients' blindness. Moreover, it only included 15 patients, of whom five were lost at 3 weeks and whose data were not analyzed. The effect of potential co-interventions was not analyzed either. All these shortcomings imply that this

study has a high risk of having been affected by multiple biases (Tables 1 and 2) [18], making it impossible to rule out the possibility that ultrasound may simply be acting as a placebo for LBP patients. In fact, this is consistent with results from previous systematic reviews on the effect of ultrasound for other musculoskeletal problems [30,31]. Another study, which also had a high risk of bias, compared ultrasound with spinal manipulation. In LBP patients, spinal manipulation has shown to lead to reductions in pain and disability that are either very small or nonexistent [5,32–35], and which might be attributed to placebo and other unspecific effects, because they are similar to those from a sham procedure [36]. Therefore, results from this study, which show that ultrasound is even less effective [16], cannot be seen as supporting its clinical use.

There is no evidence on the efficacy of shock wave when compared with a sham procedure. Therefore, it is impossible to rule out that any clinical effect attributed to this procedure is actually because of unspecific effects (eg, placebo, Hawthorne, etc). The only study on this procedure has found that, at 3 weeks, it is “not inferior” to TENS for patients with LBP and no leg pain (Table 3) [15]. However, as acknowledged by the authors of the original study [15], the evidence on the efficacy of TENS for LBP patients is limited and inconsistent [37], and therefore, results from this study cannot be seen as supporting the clinical use of shock wave therapy in routine practice [15].

Most differences in clinical evolution, which were found in the studies included in this review, are so small that they can be seen as clinically irrelevant [38,39]. The only exception is the evolution of pain at 3 weeks when comparing ultrasound with spinal manipulation, which favors the latter [16].

Publication bias generally favors positive results. Therefore, potential publication bias is not likely to affect the negative conclusions from this review.

As opposed to the study on shock wave [15], the quality of reporting of the studies on ultrasound was very poor [16–18]. Future studies in this area should follow the CONSORT recommendations on reporting for trials on nonpharmacologic interventions [40].

More than 200 treatment modalities are currently available for LBP patients, including many nonpharmacologic therapies. Most clinically and commercially available procedures have never shown to be effective, safe, and cost-effective, which raises ethical, clinical, and economic concerns [41]. The decision to finance or continue to finance a treatment should be based on solid evidence deriving from high-quality RCTs on its efficacy versus a sham procedure, and its effectiveness and cost-effectiveness versus other forms of treatment, as well as on results from post-marketing surveillance [41]. This should be particularly so in times of cost containment.

Conducting low-quality RCTs implies wasting effort, time, and resources that might have been used more

efficiently. Similarly, comparing the effectiveness of procedures that have not previously shown to be better than placebo (or the corresponding sham procedure) may be useless because it is impossible to rule out that such a study would only compare the relative size of their placebo effects [41].

Therefore, sound studies comparing ultrasound and shock wave with appropriate sham procedures should be conducted before comparing them with other forms of treatment for LBP patients. These other forms of treatment should also have previously shown to be effective. Moreover, further studies should ensure a clear and standardized definition of the type of patients to be included, with detailed inclusion and exclusion criteria; standardized treatment protocols; masking of the randomization procedure, patients, treatment providers, outcome assessment, and statistical analysis; use of standardized instruments to measure clinically relevant outcomes; samples that are large enough to ensure statistical power to detect clinically meaningful differences; implementation of measures to minimize losses to follow-up and crossover; “intention-to-treat” and “as-treated” analyses; systematic report of co-interventions and adverse events; assessment of costs and cost-effectiveness across treatment groups; and quality of reporting.

In conclusion, the main finding of this systematic review is that the available trials do not demonstrate the efficacy or effectiveness of ultrasound and shock wave for treating LBP, whether acute or chronic, with or without leg pain. This suggests that the resources assigned to fund these forms of treatment may be wasted and that the burden caused to patients (eg, visits to practitioners, time, costs, etc), may be unjustified. As a result, these procedures should only be used in the context of further research focused on assessing their efficacy, effectiveness, and cost-effectiveness, and their clinical use should be discouraged.

Acknowledgments

The authors are grateful to Dr Mohammad A. Mohseni-Bandpei [16] for having provided the correct data corresponding to a table that was misprinted in the publication and to him, Ms Barker [15], Dr Jeremy Fairbank [15], and Dr Nouredin Nakhostin Ansari [18] for having provided additional data on their studies.

This study was jointly funded by the Spanish Health Ministry's Instituto de Salud Carlos III (FIS 03/0908), and the Kovacs Foundation, an independent not for profit institution specializing in neck and back pain research, with no links to the health industry. The funding institutions had no role in the design and conduct of the study; data collection; management, analysis, and interpretation of the data; preparation, review, and approval of the manuscript; or the decision to submit the article for publication.

References

- [1] Waddell G. *The back pain revolution*. Edinburgh, UK: Churchill Livingstone, 1998.
- [2] Maxwell L. Therapeutic ultrasound: its effects on the cellular and molecular mechanisms of inflammation and repair. *Physiotherapy* 1992;78:421–6.
- [3] Sweitzer RW. Ultrasound. In: Hecox B, Weisberg J, Tsega M, editors. *Physical agents*. CT: Appleton and Lange, 1994:163–92.
- [4] Airaksinen O, Brox JI, Cedraschi C, et al. European guidelines for the management of chronic nonspecific low back pain. *Eur Spine J* 2006;15(Supplement 2):S192–300.
- [5] Chou R, Huffman LH, American Pain Society, American College of Physicians. Nonpharmacologic therapies for acute and chronic low back pain: a review of the evidence for an American Pain Society/American College of Physicians clinical practice guideline. *Ann Intern Med* 2007;147:492–504.
- [6] van Tulder MW, Becker A, Bekkering T, et al., COST B13 working group on guidelines for the management of acute low back pain in primary care. European guidelines for the management of acute nonspecific low back pain in primary care. *Eur Spine J* 2006;15(Supplement 2):S169–91.
- [7] Nussbaum EL. Ultrasound: to heat or not to heat—that is the question. *Phys Ther Rev* 1997;2:59–72.
- [8] Roebroec ME, Dekker J, Oostendorp RA. The use of therapeutic ultrasound by physical therapists in Dutch primary health care. *Phys Ther* 1998;78:470–8.
- [9] Robertson VJ, Spurr D. Electrophysical agents: implications of EPA availability and use in undergraduate clinical placements. *Physiotherapy* 1998;84:335–44.
- [10] Lindsay DM, Dearnness J, McGinley CC. Electrotherapy usage trends in private physiotherapy practice in Alberta. *Physiother Can* 1995;47:30–4.
- [11] Di Iorio D, Henley E, Doughty A. A survey of primary care physician practice patterns and adherence to acute low back problem guidelines. *Arch Fam Med* 2000;9:1015–21.
- [12] Rutjes AW, Nuesch E, Sterchi R, et al. Therapeutic ultrasound for osteoarthritis of the knee or hip. *Cochrane Database Syst Rev* 2010;(1). CD003132.
- [13] Brosseau L, Tugwell P, Wells GA, et al., Philadelphia Panel (Canada and the United States of America). Philadelphia Panel evidence-based clinical practice guidelines on selected rehabilitation interventions for low back pain [with systematic review]. *Phys Ther* 2001;81:1641–74.
- [14] Furlan AD, Pennick V, Bombardier C, van Tulder M, Editorial Board of the Cochrane Back Review Group. 2009 updated method guidelines for systematic reviews in the Cochrane Back Review Group. *Spine* 2009;34:1929–41.
- [15] Barker KL, Elliott CJ, Sackley CM, et al. Treatment of chronic back pain by sensory discrimination training. A Phase I RCT of a novel device (FairMed) vs. TENS. *BMC Musculoskelet Disord* 2008;9:97.
- [16] Mohseni-Bandpei M, Critchley J, Staunton T, et al. A prospective randomized controlled trial of spinal manipulation and ultrasound in the treatment of chronic low back pain. *Physiotherapy* 2006;92:34–42.
- [17] Unlu Z, Tasci S, Tarhan S, et al. Comparison of 3 physical therapy modalities for acute pain in lumbar disc herniation measured by clinical evaluation and magnetic resonance imaging. *J Manipulative Physiol Ther* 2008;31:191–8.
- [18] Ansari NN, Ebadi S, Talebian S, et al. A randomized, single blind placebo controlled clinical trial on the effect of continuous ultrasound on low back pain. *Electromyogr Clin Neurophysiol* 2006;46:329–36.
- [19] Faas A, Chavannes AW, van Eijk JT, et al. A randomized, placebo-controlled trial of exercise therapy in patients with acute low back pain. *Spine* 1993;18:1388–95.
- [20] Nwuga VC. Ultrasound in treatment of back pain resulting from prolapsed intervertebral disc. *Arch Phys Med Rehabil* 1983;64:88–9.

- [21] Roman MP. A clinical evaluation of ultrasound by use of a placebo technic. *Phys Ther Rev* 1960;40:649–52.
- [22] Koes BW, Bouter LM, van Mameren H, et al. Randomised clinical trial of manipulative therapy and physiotherapy for persistent back and neck complaints: results of one year follow up. *BMJ* 1992;304:601–5.
- [23] Koes BW, Bouter LM, van Mameren H, et al. A blinded randomized clinical trial of manual therapy and physiotherapy for chronic back and neck complaints: physical outcome measures. *J Manipulative Physiol Ther* 1992;1:16–23.
- [24] Hurwitz EL, Morgenstern H, Harber P, et al. The effectiveness of physical modalities among patients with low back pain randomized to chiropractic care: findings from the UCLA low back pain study. *J Manipulative Physiol Ther* 2002;25:10–20.
- [25] Borman P, Keskin D, Bodur H. The efficacy of lumbar traction in the management of patients with low back pain. *Rheumatol Int* 2003;23:82–6.
- [26] Fernández-de-las-Peñas C, Fernández-Carnero J, Palomeque del cerro L, Miangolarra-Page JC. Manipulative treatment vs. conventional physiotherapy treatment in whiplash injury: a randomized controlled trial. *J Whiplash and Related Disorders* 2004;3:73–90.
- [27] Koldas Dogan S, Sonel Tur B, Kurtais Y, Atay MB. Comparison of three different approaches in the treatment of chronic low back pain. *Clin Rheumatol* 2008;27:873–81.
- [28] Clarke J, van Tulder M, Blomberg S, et al. Traction for low-back pain with or without sciatica. *Cochrane Database Syst Rev* 2007;(2). CD003010.
- [29] Yousefi-Nooraie R, Schonstein E, Heidari K, et al. Low level laser therapy for nonspecific low-back pain. *Cochrane Database Syst Rev* 2008;(2). CD005107.
- [30] Robertson VJ, Baker KG. A review of therapeutic ultrasound: effectiveness studies. *Phys Ther* 2001;81:133–50.
- [31] van der Windt DA, van der Heijden GJ, van den Berg SG, et al. Ultrasound therapy for musculoskeletal disorders. *Pain* 1999;81:257–71.
- [32] Cherkin DC, Sherman KJ, Deyo RA, et al. A review of the evidence for the effectiveness, safety, and cost of acupuncture, massage therapy, and spinal manipulation for back pain. *Ann Intern Med* 2003;138:898–906.
- [33] Eisenberg DM, Post DE, Davis RB, et al. Addition of choice of complementary therapies to usual care for acute low back pain. A randomized controlled trial. *Spine* 2007;32:151–8.
- [34] Ferreira ML, Ferreira PH, Latimer J, et al. Does spinal manipulative therapy help people with chronic low back pain? *Aust J Physiother* 2002;48:277–84.
- [35] UK BEAM Trial Team. United Kingdom back pain exercise and manipulation (UK BEAM) randomized trial: effectiveness of physical treatments for back pain in primary care. *BMJ* 2004;329:1377.
- [36] Licciardone JCC, Stoll ST, Fulda KG, et al. Osteopathic manipulative treatment for chronic low back pain. A randomized controlled trial. *Spine* 2003;28:1355–62.
- [37] Khadilkar A, Milne S, Brosseau L, et al. Transcutaneous electrical nerve stimulation for the treatment of chronic low back pain: a systematic review. *Spine* 2005;30:2657–66.
- [38] Kovacs FM, Abaira V, Royuela A, et al. Minimal clinically important change for pain intensity and disability in patients with nonspecific low back pain. *Spine* 2007;32:2915–20.
- [39] Ostelo RW, Deyo RA, Stratford P, et al. Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. *Spine* 2008;33:90–4.
- [40] Boutron I, Moher D, Tugwell P, et al. A checklist to evaluate a report of a nonpharmacological trial (CLEAR NPT) was developed using consensus. *J Clin Epidemiol* 2005;58:1233–40.
- [41] Carragee EJ, Deyo RA, Kovacs FM, et al. Clinical research. Is the spine field a mine field? *Spine* 2009;34:423–30.
- [42] Waterfield J, Sim J. Clinical assessment of pain by visual analogue scale. *Br J Ther Rehabil* 1996;3:94–7.
- [43] Fairbank J, Couper J, Davies J, et al. The Oswestry low back pain disability questionnaire. *Physiotherapy* 1980;66:271–3.
- [44] William R, Blinkley J, Bloch R, et al. Reliability of the modified-modified Schober and double inclinometer methods for measuring lumbar flexion and extension. *Phys Ther* 1993;73:33–44.
- [45] Biering-Sorensen F. Physical measurements as risk indicators for low back trouble over a one year period. *Spine* 1984;9:106–19.
- [46] Jensen MP, Karoly P, Braver S. The measurement of clinical pain intensity: a comparison of six methods. *Pain* 1986;27:117–26.
- [47] Fairbank J, Pynsent P. The Oswestry Disability Index. Available at: http://www.chirogeek.com/001_Oswestry_20.htm. Accessed: August 31, 2010.
- [48] Harding VR, Williams AC, Richardson PH, et al. The development of a battery of measures for assessing physical functioning of chronic pain patients. *Pain* 1994;58:367–75.
- [49] Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand* 1983;67:361–70.
- [50] Vlaeyen JWS, Kole-Snijders AMJ, Rotteveel AM, et al. The role of fear of movement/(re)injury in pain disability. *J Occup Rehabil* 1995;5:235–52.
- [51] Sullivan MJL, Bishop SC, Pivik J. The Pain Catastrophizing Scale: development and validation. *Psychol Assess* 1995;7:524–32.
- [52] Nicholas MK. The pain self-efficacy questionnaire: taking pain into account. *Eur J Pain* 2007;11:153–63.
- [53] Guy W. ECDEU assessment manual for psychopharmacology. DHEW Publication No. ADM 76–338. Washington, DC: US Government Printing Office, 1976.

Appendix

Search strategy used for all reviews

MEDLINE (PubMed)

#2 "Low Back Pain"[Mesh]

#3 back pain[tw] OR lumbago[tw] OR backache[tw] OR spine pain[tw]

#4 #2 or #3 #7 "Electric Stimulation Therapy"[Mesh] NOT "Transcutaneous Electric Nerve Stimulation"[Majr] #9 electro stimulation[tiab] OR electrotherapy[tiab] OR electro therapy[tiab] OR short wave*[tiab] OR microwave*[tiab] OR ultrasound*[tiab] OR Laser*[tiab] OR magnetotherapy[tiab] OR pulsed radiofrequenc*[tiab] OR electro-acupuncture [tiab] OR electroacupuncture[tiab] OR electrostimulation [tiab] OR electrothermal[tiab] OR neuromodulation[tiab] OR intramuscular stimulation[tiab]

#10 #7 or #9

#11 transcutaneous electric nerve stimulation[ti] OR TENS[ti]

#12 #10 not #11

#13 #4 AND #12

#14 (randomized controlled trial[pt] OR controlled clinical trial[pt] OR randomized[tiab] OR placebo[tiab] OR drug therapy[sh] OR randomly[tiab] OR trial[tiab] OR

groups[tiab] NOT (animals[mh] NOT (humans[mh] AND animals[mh]))

#15 #13 AND #14

#16 ("1"[Publication Date] : "2009/12/31"[Publication Date]) AND (#13 AND #14)

#17 (#15) NOT #16

CENTRAL (The Cochrane Library)

#1 MeSH descriptor Low Back Pain explode all trees

#2 back pain OR lumbago OR backache OR spine pain

#3 (#1 OR #2)

#4 MeSH descriptor Electric Stimulation Therapy explode all trees #5 electric stimulation OR electrotherapy OR electro therapy OR short wave* OR microwave* OR ultrasound* OR Laser* OR magnetotherapy OR pulsed radiofrequenc* OR electro-acupuncture OR electroacupuncture OR electrostimulation OR electrothermal OR neuromodulation OR intramuscular stimulation

#6 (#4 OR #5)

#7 transcutaneous electric nerve stimulation:ti OR TENS:ti

#8 (#6 AND NOT #7)

#9 (#3 AND #8)

#10 (#9), from 1900 to 2009

#11 (#9 AND NOT #10)

EMBASE (Ovid)

1 exp Backache/

2 exp Low Back Pain/

3 back pain.mp.

4 back.ti.

5 lumbago.mp.

6 spine pain.mp.

7 1 or 2 or 3 or 4 or 5 or 6

8 exp electrostimulation therapy/

9 (electric stimulation or electrotherapy or electro therapy or short wave* or microwave* or ultrasound* or Laser* or magnetotherapy or pulsed radiofrequenc* or electroacupuncture or electroacupuncture or electrostimulation or electrothermal or neuromodulation or intramuscular stimulation).ti,ab.

10 8 or 9

11 (transcutaneous electric nerve stimulation or TENS).ti.

12 10 not 11

13 7 and 12

14 random:.tw. or clinical trial:.mp. or exp health care quality/

15 13 and 14

16 limit 15 to yr="1974 - 2009"

17 15 not 16



Artículo especial

Análisis de la investigación española sobre dolencias del cuello y la espalda (1992–2006)

Analysis of Spanish research about neck and back complaints (1992–2006)

Gerard Urrútia^{a,b,*} y Xavier Bonfill^{a,b}

^a Servicio de Epidemiología Clínica y Salud Pública, Centro Cochrane Iberoamericano, Hospital de la Santa Creu i Sant Pau (UAB), Barcelona, España

^b CIBER Epidemiología y Salud Pública (CIBERESP), Barcelona, España

INFORMACIÓN DEL ARTÍCULO

Historia del artículo:

Recibido el 16 de julio de 2009

Aceptado el 29 de octubre de 2009

On-line el 30 de diciembre de 2009

Introducción

Pocos estudios han analizado la productividad y la eficiencia de la inversión española en investigación médica. Para ese análisis, los síndromes mecánicos del raquis (SMR) constituyen un campo adecuado dada su frecuencia, el coste, la variabilidad y el escaso fundamento científico de su tratamiento clínico y el hecho de implicar a múltiples profesiones sanitarias (Medicina, Psicología, Fisioterapia, Enfermería, etc.) y especialidades médicas (Atención Primaria, Rehabilitación, Radiología, Reumatología, Cirugía Ortopédica, Neurocirugía, etc.)^{1,2}.

El objetivo de este estudio fue describir la investigación española sobre SMR, su importe, las fuentes de financiación y la productividad.

Material y métodos

Criterios de inclusión

Se incluyeron los proyectos de investigación que: a) se refirieran a SMR a cualquier nivel del raquis y no incluyeran pacientes con otras enfermedades —inflamatorias, neoplásicas, etc.—; b) usaran un método de investigación formal; c) al menos un autor trabajara en España, y d) se hubieran presentado para su financiación, realizados o publicados entre el 1 de enero de 1992 y el 31 de diciembre de 2006.

Identificación de los proyectos, los investigadores y las entidades financiadoras

Se usaron estrategias de búsqueda electrónica en bases de datos de estudios publicados (Medline y Embase) y en curso (Clinicaltrials.gov y Current Controlled Trials).

Se contactó con los investigadores, los servicios hospitalarios y las entidades financiadoras mencionados en los proyectos hallados, con los investigadores y las entidades que esas fuentes identificaron, con el Fondo de Investigación Sanitaria (FIS) y las agencias de evaluación de tecnologías sanitarias (AETS) españolas y con los servicios de cirugía ortopédica y traumatología, neurocirugía, reumatología, rehabilitación y las unidades del dolor de los 275 hospitales públicos generales identificados en la base de datos interna del Centro Cochrane Iberoamericano. Dado su gran número, se descartó dirigirse a las gerencias de Atención Primaria existentes en España.

En una segunda ronda, se envió un recordatorio a quienes no habían respondido a la primera. En la tercera ronda, se llamó por teléfono a quienes no habían respondido a las 2 rondas previas.

Se realizaron 87 comunicaciones escritas y 39 comunicaciones telefónicas a los 52 investigadores identificados, 2.375 comunicaciones escritas a 1.100 servicios de 281 hospitales y 22 comunicaciones escritas y 14 llamadas telefónicas a las 18 entidades financiadoras identificadas (tabla 1).

Selección de los proyectos

Se introdujeron los proyectos identificados en una base de datos y se eliminaron las redundancias. Los autores valoraron de manera independiente la inclusión de los proyectos en función

* Autor para correspondencia.

◊ Gerard Urrútia pertenece al Programa de Doctorado en Salud Pública y Metodología de la Investigación de la Universitat Autònoma de Barcelona.

Correo electrónico: gurrutia@santpau.cat (G. Urrútia).

Tabla 1
Proyectos sobre síndromes mecánicos del raquis evaluados y financiados por las distintas entidades, recursos destinados y estado de los proyectos. Todos los datos económicos se expresan en € constantes de 2007*

Entidad financiadora	Total de proyectos F y E, n	Estado de los proyectos			
		Publicados ^a n=80; FIT=115,30		En curso n=34	Improductivos n=12
		n/importe, €	Referencias y FIT	n/importe, €	n/importe, €
Fundación Kovacs	F: 51 E: 51	19/7.388.535	63-82 FIT=49,12	32/4.500.054	0
FIS del Instituto de Salud Carlos III	F: 18 E: 33	6/316.361	52,57-59,64,65,76 FIT=7,793	5/154.072	6/395.814
AETS del Instituto de Salud Carlos III	F: 6 E: NF	1/3.361	53,78 FIT=2,351	3/10.083	1/NF
AETSA	F: 5 E: NF	1/10.449	54,60 FIT=6,294	1/16.110	2/22.881
Osteba	F: 4 E: NF	1/22.200	83 FIT=0,000	0	3/47.651
Cons IB	E: 2 E: NF	2/38.781	62,72 FIT=2,026	0	0
ISSD Mallorca	E: 2 E: NF	0	0	2/6.834	0
IMSERSO	F: 2 E: NF	0	0	2/36.994	0
AATRM	F: 1 E: NF	1/30.318	61 FIT=1,824	0	0
Cons Ed IB	F: 1 E: NF	1/2.495	66 FIT=1,000	0	0
Cons CL- ICSCCL	F: 1 E: NF	0	0	1/9.747	0
Cons CM	F: 1 E: NF	1/ NF	55 FIT=0,000		
GESA	F: 1 E: NF	1/120.420	72 FIT=4,556	0	0
CCIB	F: 1 E: 1	1/3.330	80 FIT=4,654	0	0
Bayer	F: 1 E: NF	1/NF	56 FIT=0,825	0	0
ASAC	F: 1 E: NF	1/NF	54 FIT=1,458	0	0
GSK	F: 1 E: NF	1/NF	52 FIT=1,327	0	0
UAB	F: 2 E: NF	2/NF	53,56 FIT=3,176	0	0
No requirieron financiación	8	8	4-11 FIT=15,16	0	0
Financiación de origen desconocido	40	49/NF	12-51 FIT=29,82	0	0

AATRM: Agència d'Avaluació de Tecnologia i Recerca Mèdiques de Catalunya; AETS: Agencia de Evaluación de Tecnología Sanitaria; AETSA: Agencia de Evaluación de Tecnologías Sanitarias de Andalucía; ASAC: ASAC Pharmaceutical International; CCIB: Centro Cochrane Iberoamericano; Cons CL: Consejería de Sanidad de Castilla-León; Cons CM: Consejería de Sanidad de Castilla-La Mancha; Cons Ed IB: Consejería de Educación de las Islas Baleares; Cons IB: Consejería de Sanidad de las Islas Baleares; E: evaluado; F: financiado; FIS: Fondo de Investigaciones Sanitarias; FIT: factor de impacto total (acumulado); GESA: Gas y Electricidad Sociedad Anónima; GSK: Glaxo-Smith-Kline; ICSCCL: Instituto de Ciencias de la Salud de Castilla-León; IMSERSO: Instituto de Migraciones y Servicios Sociales; ISSD Mallorca: Instituto de Servicios Sociales y Deportivos de Mallorca; NF: datos no facilitados por la entidad financiadora ni por los investigadores; Osteba: Servicio de Evaluación de Tecnologías Sanitarias del País Vasco; UAB: Universidad Autónoma de Barcelona.

* Todos los datos económicos se expresan en € constantes de 2007.

^a En los proyectos cofinanciados por varias entidades, la publicación y el factor de impacto correspondiente se han asignado enteramente a cada una de ellas. Por eso, la suma del número de proyectos productivos, publicaciones y factor de impacto de todas las entidades es mayor que el total. Las entidades que compartieron la financiación de cada estudio se muestran en la tabla 2.

de la información incluida en el artículo o el registro correspondiente. Sus valoraciones concordaron.

Variables

Se pidió a cada entidad financiadora el título de los proyectos sobre SMR que les presentaron entre el 1 de enero de 1992 y el 31 de diciembre de 2006, el nombre y los datos del investigador principal y el importe solicitado y concedido (0 en los denegados).

Para cada proyecto, se pidió a los investigadores título, tipo (“«básico»” o “«clínico o en servicios de salud»”, y se distinguió

“«ensayo clínico aleatorizado»” [ECA], “«revisión sistemática»” u “«otros»”), importe solicitado y concedido por cada entidad financiadora y estado (“«denegado», “«en curso»”, “«en prensa»” o “«publicado»”). Además, se les envió una encuesta sobre la convocatoria y la labor de las entidades financiadoras.

Análisis

Para cada proyecto se cotejó la información sobre el importe solicitado y concedido que aportaron los investigadores y las entidades financiadoras. Los datos concordaron.

Se contaron los proyectos no incluidos en la lista aportada por las entidades financiadoras y en los que el investigador no desveló el origen de los fondos ni figuró en el artículo o el registro correspondiente (“«estudios con financiación de origen no declarado»”) y el número de proyectos en el que ni las entidades financiadoras ni los investigadores desvelaron su importe (“«estudios con financiación de cuantía no declarada»”).

Se asignó a cada proyecto el factor de impacto (FI) de la revista en el año en el que se publicó³, y se computó el de 1999 a los publicados antes de ese año. Se calculó el FI acumulado de cada entidad mediante la suma de sus publicaciones (FI total [FIT]). Para las entidades que habían publicado 2 o más artículos, se calculó su media (media del FI). El FI de un estudio se asignó a

todas las entidades que lo cofinanciaron. Para cada entidad financiadora se calculó por separado el FI de los proyectos que cofinanció y los que financió en solitario.

Se consideraron “«proyectos denegados que se publicaron»” aquellos que una entidad financió después de que otra los rechazara, y se publicaron antes del 31 de diciembre de 2008. Se calculó su número, porcentaje y FI. El FI de los estudios publicados entre el 31 de diciembre de 2006 y el 31 de diciembre de 2008 no se usó para calcular el de las publicaciones realizadas entre 1992 y 2006.

Se identificaron los proyectos finalizados antes del 31 de diciembre de 2004 y no aceptados por ninguna revista científica antes del 31 de diciembre de 2006 (“«proyectos improductivos»”). Para cada entidad se calculó su número y porcentaje.

Tabla 2

Importe destinado por cada entidad financiadora a los distintos tipos de proyectos en los que se pudo identificar la fuente de financiación*

Entidad y n de proyectos financiados (total=81)*	Estudios básicos (n/importe)	Estudios clínicos** (n/importe)			Total invertido declarado por la entidad, €
		ECA (n=16)	RS (n=9)	Otros diseños (n=98)	
Fundación Kovacs, 51 proyectos	0	13 ^a /4.516.874	8 ^b /247.346	30 ^c /7.124.369	11.888.589
FIS del Instituto de Salud Carlos III, 18 proyectos	2/275.383	4 ^d /350.119	3 ^e /15.838	8 ^f /224.908	866.248
AETS del Instituto de Salud Carlos III, 6 proyectos	0	0	4 ^h /13.444	1 ^g /NF 2 ⁱ /NF	13.444
AETSA, 5 proyectos	1/18.840	1/10.449 1 ^j /NF	0	2/20.151	49.440
Osteba, 4 proyectos	0	0	0	4/69.851	69.851
Cons IB, 4 proyectos	0	0	0	2 ^k /38.781	38.781
ISSD Mallorca, 2 proyectos	0	2 ^l /6.834	0	0	6.834
IMERSO, 2 proyectos	0	0	0	2 ^m /36.994	36.994
AATRM, un proyecto	0	0	0	1/30.318	30.318
Cons Ed IB, un proyecto	0	0	0	1 ⁿ /2.495	2.495
Cons CL-ICSCCL, un proyecto	0	1 ^o /9.747	0	0	9.747
Cons CM, un proyecto	0	0	0	1	NF
GESA, un proyecto	0	0	0	1 ^p /120.420	120.420
CCIB, un proyecto	0	0	1/3.330 €	0	3.330
Bayer, un proyecto	0	0	0	1 ^q /NF	NF
ASAC, un proyecto	0	1 ^r /NF	0	0	NF
GSK, un proyecto	0	0	0	1 ^s /NF	NF
UAB, 2 proyectos	0	0	0	2 ^t /NF	NF
Total	294.223	4.894.023	279.958	7.668.287	13.136.491

AATRM: Agència d'Avaluació de Tecnologia i Recerca Mèdiques de Catalunya; AETS: Agencia de Evaluación de Tecnología Sanitaria; AETSA: Agencia de Evaluación de Tecnologías Sanitarias de Andalucía; ASAC: ASAC Pharmaceutical International; CCIB: Centro Cochrane Iberoamericano; Cons CL: Consejería de Sanidad de Castilla-León; Cons CM: Consejería de Sanidad de Castilla-La Mancha; Cons Ed IB: Consejería de Educación de las Islas Baleares; Cons IB: Consejería de Sanidad de las Islas Baleares; ECA: ensayo clínico aleatorizado; FIS: Fondo de Investigaciones Sanitarias; GESA: Gas y Electricidad Sociedad Anónima; GSK: Glaxo-Smith-Kline; ICSCCL: Instituto de Ciencias de la Salud de Castilla-León; IMERSO: Instituto de Migraciones y Servicios Sociales; ISSD Mallorca: Instituto de Servicios Sociales y Deportivos de Mallorca; NF: datos no facilitados por la entidad financiadora ni por los investigadores; Osteba: Servicio de Evaluación de Tecnologías Sanitarias del País Vasco; RS: revisión sistemática; UAB: Universidad Autónoma de Barcelona.

* Todos los datos económicos se expresan en € constantes de 2007.

** Varios proyectos se cofinanciaron por más de una entidad, por lo que la suma del número de proyectos financiados por cada una (105) es mayor que el número total de proyectos (81).

^a Tres cofinanciados con el FIS, 2 con el ISSD Mallorca, 2 con el IMERSO y uno con la Cons CL-ICSCCL. El resto en solitario.

^b Tres cofinanciadas con el FIS, 4 con la AETS y otro con entidades extranjeras (no incluidas en este estudio).

^c Dos cofinanciados con el FIS, 3 con la Consejería de Salud de Baleares, uno con la Cons Ed IB y otro con GESA. El resto en solitario.

^d Tres cofinanciados con la Fundación Kovacs y uno financiado en solitario.

^e Las 3 cofinanciadas con la Fundación Kovacs.

^f Dos cofinanciados con la Fundación Kovacs y el resto en solitario.

^g Cofinanciado con GSK.

^h Cofinanciados con la Fundación Kovacs.

ⁱ Uno cofinanciado con la UAB y el otro en solitario.

^j Cofinanciado con ASAC.

^k Cofinanciados con la Fundación Kovacs.

^l Cofinanciados con la Fundación Kovacs.

^m Cofinanciados con la Fundación Kovacs.

ⁿ Cofinanciado con la Fundación Kovacs.

^o Cofinanciado con la Fundación Kovacs.

^p Cofinanciado con la Fundación Kovacs.

^q Cofinanciado con la UAB.

^r Cofinanciado con la AETSA.

^s Cofinanciado con el FIS del Instituto de Salud Carlos III.

^t Uno cofinanciado con Bayer y otro con la AETS del Instituto de Salud Carlos III.

Los importes se convirtieron a € constantes de 2007, y se aplicaron los aumentos del IPC publicados para cada año por el Instituto Nacional de Estadística (disponible en: <http://www.ine.es/>). Se consideró «inversión improductiva» la destinada a proyectos improductivos.

Resultados

Proyectos

La búsqueda electrónica identificó 707 referencias. Eliminadas las redundancias, se identificaron 113 proyectos de investigación, 41 investigadores y 17 entidades financiadoras. Los contactos con investigadores, hospitales y servicios permitieron identificar

adicionalmente 16 proyectos publicados en revistas no indexadas, 11 investigadores y una entidad (tablas 1 y 2).

Diez investigadores (19,2%) no pudieron localizarse, 28 (53,8%) no respondieron, uno (1,9%) se negó explícitamente a aportar datos y 13 (25%) respondieron a los cuestionarios enviados. No aportó datos ningún autor de proyectos improductivos, estudios retrospectivos o con financiación de origen o cuantía no declaradas. Los 13 autores que sí lo hicieron eran responsables de 75 (58,1%) de los 129 proyectos. Ninguno había sido improductivo, todos contaban con financiación transparente e incluían todas las revisiones sistemáticas y 15 (93,8%) de los 16 ECA.

Respondieron 79 servicios de 55 hospitales. Diecisiete (21,5%) habían participado en estudios sobre SMR.

Tabla 3

Indicadores de la calidad del proceso de evaluación de las entidades que han evaluado algún proyecto ya publicado

Entidad	Total de proyectos publicados, FIT, FIM (DE) ^a	Publicaciones, FIT, FIM (DE) de los estudios cofinanciados con terceros ^a	Publicaciones, FIT, FIM (DE) de los estudios financiados en solitario ^a	Publicaciones, FIT, FIM (DE) del FI de los estudios denegados ^{a,b}
Fundación Kovacs	19 FIT=49,120 FIM=2,729 (4,02)	7 FIT=9,122 FIM=1,523 (1,59)	12 FIT=39,998 FIM=3,333 (4,7)	0
FIS del Instituto de Salud Carlos III	7 FIT=7,793 FIM=1,113 (0,57)	4 FIT=3,741 FIM=1,267 (0,39)	3 FIT=4,052 FIM=0,908 (0,80)	7 FIT=16,190 FIM=2,313 (0,48)
AETS del Instituto de Salud Carlos III	2 FIT=2,351 FIM=1,175 (1,66)	2 FIT=2,351 FIM=1,175 (1,66)	0	NF
AETSA	2 FIT=6,294 FIM=3,147 (2,38)	1 FIT=1,458	1 FIT=4,836	NF
Osteba	1 FIT=0,00	0	1 FIT=0,00	NF
Cons IB	2 FIT=2,026 FIM=0,506 (0,60)	1 FIT=0,825	1 FIM=1,201	NF
AATRM	1 FIT=1,824	0	1 FIT=1,824	NF
Cons Ed IB	1 FIT=1,000	1 FIT=1,000	0	NF
GESA	1 FIT=4,556	1 FIT=4,556	0	NF
CClb	1 FIT=4,654	0	1 FI=4,654	0
Bayer	1 FIT=0,825	1 FIT=0,825	0	NF
ASAC	1 FIT=1,458	1 FIT=1,458	0	NF
GSK	1 FIT=1,327	1 FI=1,327	0	NF
UAB	2 FIT=3,176 FIM=1,588 (1,07)	2 FIT=3,176 FIM=1,588 (1,07)	0	NF
Cons CM	1 FI=0,000	0	1 FI=0,000	NF

AATRM: Agència d'Avaluació de Tecnologia i Recerca Mèdiques de Catalunya; AETS: Agencia de Evaluación de Tecnología Sanitaria; AETSA: Agencia de Evaluación de Tecnologías Sanitarias de Andalucía; ASAC: ASAC Pharmaceutical International; DE: desviación estándar; CClb: Centro Cochrane Iberoamericano; Cons CM: Consejería de Sanidad de Castilla-La Mancha; Cons Ed IB: Consejería de Educación de las Islas Baleares; Cons IB: Consejería de Sanidad de las Islas Baleares; FI: factor de impacto; FIM: media del factor de impacto; FIS: Fondo de Investigaciones Sanitarias; FIT: factor de impacto total; GESA: Gas y Electricidad Sociedad Anónima; GSK: Glaxo-Smith-Kline; NF: datos no facilitados por la entidad financiadora ni por los investigadores; Osteba: Servicio de Evaluación de Tecnologías Sanitarias del País Vasco; UAB: Universidad Autónoma de Barcelona.

^a FIT acumulado (sumatoria de las publicaciones).

^b Solo la Fundación Kovacs y el CClb informaron sobre el número de proyectos que denegaron, pero los investigadores identificaron 15 proyectos denegados por el FIS.

Respondieron 12 entidades financiadoras. Las dependientes del Instituto Carlos III (ISCIII) —el FIS y la AETS— solo aportaron datos de los proyectos aprobados (no denegados) durante 2000–2004. La Fundación Kovacs (FK) declaró no haber denegado ningún proyecto. Las demás entidades no aportaron datos sobre los proyectos que denegaron. Los investigadores identificaron 15 proyectos denegados por el FIS y 2 financiados en años sobre los que esta entidad no facilitó datos.

Tipo de investigación

El FIS y la AETS de Andalucía (AETSA) financiaron los 3 proyectos de investigación básica. Todos fueron improductivos. De los 126 clínicos, 16 fueron ECA y 9 fueron revisiones sistemáticas (tabla 2).

Financiación

Ocho proyectos (6,2%) se realizaron sin financiación^{4–11}. No se desveló el origen de la financiación de 40 (31,0%)^{12–51} ni el importe de 6 (3,9%), de los que uno resultó improductivo (tabla 1)^{52–56}. Los 75 que se financiaron transparentemente costaron 13.136.490 € y 30 de ellos ya se habían publicado^{57–83}. Los mayores inversores fueron la FK (el 90,5% de la inversión total conocida) y el FIS (6,6%) (tabla 2).

Calidad del proceso de evaluación

Entre las entidades que evaluaron más de 10 proyectos, el FIT de la FK fue de 49,120 y el del FIS de 7,793 (tabla 3).

De los 15 proyectos rechazados por el FIS, la FK financió 12. Siete se habían publicado con un FIT de 16,190 (tabla 3)^{68,74,75,84–87}.

El FIS, la AETS del País Vasco (Servicio de Evaluación de Tecnologías Sanitarias del País Vasco [Osteba]), la AETSA y la AETS financiaron los 12 proyectos improductivos (6 [el 33% de los 18 que financió], 3 [75%], 2 [40%] y 1 [16,7%], respectivamente) (tabla 1).

La financiación improductiva fue para el FIS de 560.350 € (el 45,7% de su inversión), para el Osteba de 47.651 € (68,2%) y para la AETSA de 22.881 € (46,3%). La de la AETS no pudo calcularse porque la agencia no desveló el importe de varios proyectos (tablas 1 y 2).

Dado el bajo número de investigadores que valoraron a las entidades, se decidió no analizar sus respuestas.

Discusión

Entre 1992 y 2006 estuvieron activos en España 129 proyectos de investigación sobre SMR. Doce fueron improductivos, incluyendo todos los básicos, 34 seguían en curso y 83 se habían publicado (tabla 1)^{4–83}. Su calidad fue muy heterogénea y se publicaron en revistas cuyo FI oscilaba entre 0 y 18,136⁷⁰.

La financiación transparente ascendió a 13.136.490 €. La aportación de las principales entidades públicas de investigación (EPI) biomédica (ISCIII [FIS y AETS], Osteba, Agència d'Avaluació de Tecnologia i Recerca Mèdiques de Catalunya y AETSA) representó menos del 8%. La FK, una entidad privada sin ánimo de lucro ni vinculación con la industria sanitaria⁸⁸, aportó el 90,5%. Aunque se especializa en la investigación sobre SMR, esa comparación resulta abrumadora.

Globalmente, parecen deficientes los mecanismos que aplican las EPI para seleccionar los proyectos que financian, pues entre el 16,7 y el 75% de los proyectos que aprobaron resultaron improductivos. El FIS es la EPI que ha valorado más proyectos;

un tercio de los aprobados fueron improductivos y, entre los que se publicaron, el FI de los que denegó^{68,74,75,84–87} es superior al de los que aprobó (tablas 1 y 3)^{52,57–59,64,65,76}.

También es baja la eficiencia de los recursos públicos invertidos por las EPI. En el caso del FIS, de los 877.247 € invertidos, 560.350 € (45,7%) fueron improductivos. Los porcentajes de inversión improductiva son también muy altos para la AETSA (46,3%) y el Osteba (68,2%). El de la AETS no pudo calcularse porque la agencia no desveló el importe total de los fondos que destinó a este campo (tabla 1).

Resultó imposible averiguar la identidad de las entidades que financiaron aproximadamente el 30% de los proyectos (tabla 1). La mayoría de ellos correspondieron a estudios de baja calidad metodológica y tal vez no requirieron financiación externa^{12–51}.

Es loable que algunas entidades públicas y privadas hayan aportado sus datos a este estudio, aunque otras no lo han hecho (tabla 2). Con el fin de asegurar la máxima transparencia, es deseable que todas las entidades financiadoras, y especialmente las que gestionan fondos públicos, faciliten el acceso a toda la información relativa a los proyectos que deniegan y aprueban, los fondos que les destinan y las publicaciones que generan finalmente en una página web de acceso público⁸⁹.

Varias entidades financiadoras, investigadores y hospitales rechazaron aportar datos y es posible que algún proyecto de investigación pasara desapercibido. No obstante, la búsqueda fue exhaustiva e incluyó varias fuentes.

La publicación en una revista científica y su FI se han considerado como indicadores de la calidad científica de un proyecto. Aunque la precisión de esos indicadores puede ser discutida, es menos probable que artículos de baja calidad se publiquen en revistas de alto FI.

Se aceptaron los datos económicos aportados por cada entidad. Es posible que sus mecanismos contables y de auditoría sean distintos, pero a todas las entidades se les solicitaron los mismos datos y los aportados coincidieron con los facilitados por los investigadores. Además, como la inversión conjunta de todas las EPI representa una proporción mínima de la inversión (tabla 1), cualquier duda sobre la precisión o la fiabilidad de los datos que han aportado no cuestionaría el sentido general de estos resultados.

El FI de cada artículo se asignó íntegramente a todas las entidades que cofinanciaron el estudio correspondiente. No pudo fraccionarse entre las entidades cofinanciadoras proporcionalmente a su inversión porque hubo varios estudios en los que al menos una no la cuantificó (tablas 1 y 2).

El análisis se limitó a la producción científica entre 1992 y 2006, pero nada indica que antes o después de ese período se hayan dado cambios drásticos en la situación, y 15 años parece un período suficientemente largo para tener una visión ponderada de la investigación española sobre SMR. Por último, los resultados de este estudio pueden no ser generalizables a la investigación en otros campos, en los que sería deseable que se centraran futuros estudios.

En conclusión, este estudio demuestra que la inversión realizada en el ámbito de los SMR por las EPI entre 1992 y 2006 ha sido escasa e ineficiente. Además, los mecanismos con los que las EPI han seleccionado los proyectos que han financiado han sido poco válidos, pues han aprobado un porcentaje importante de proyectos improductivos y han rechazado otros de mayor calidad científica. En contraste, la entidad con mayor inversión, producción y eficiencia en este campo ha sido una entidad privada sin ánimo de lucro. Las EPI deberían incrementar los recursos económicos destinados a los SMR y, a su vez, revisar y mejorar sus mecanismos de evaluación.

Financiación

Los autores han participado en proyectos financiados por el FIS y colaboran habitualmente con las AETS españolas y la FK. Son miembros del CIBERESP, financiado por el ISCIII, y de la Red Española de Investigación en Dolencias de la Espalda, financiada por la FK.

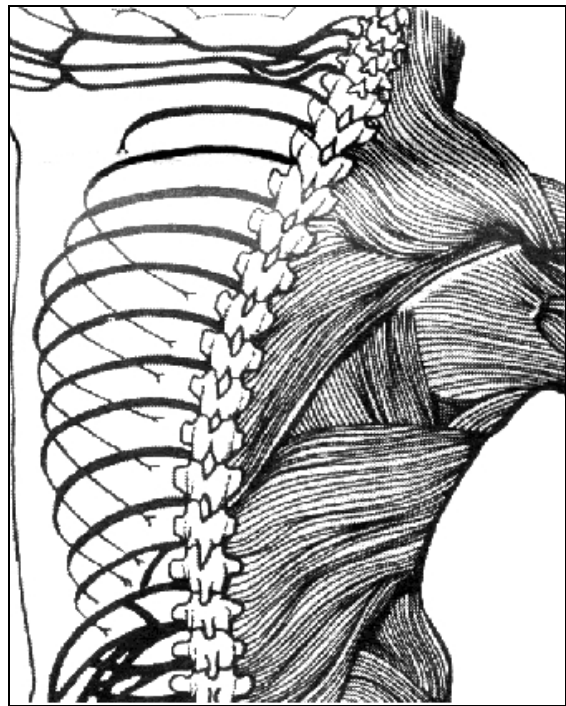
El primer autor está adscrito al Programa de Doctorado en Salud Pública y Metodología de la Investigación Biomédica de la Universitat Autònoma de Barcelona.

Bibliografía

- Waddell G. The back pain revolution. Edinburgh: Churchill Livingstone; 1998.
- Andersson GBJ. The epidemiology of spinal disorders. En: Frymoyer JW, editor. The adult spine: Principles and practice. Vol. 1. New York: Raven Press; 1991. p. 107–46.
- [consultado 14/01/2009]. Disponible en: www.accesowok.fecyt.es/news/2007/20070709.html.
- Soler T, Calderón C. The prevalence of spondylolysis in the Spanish elite athlete. *Am J Sports Med*. 2000;28:57–62.
- Méndez FJ, Gómez-Conesa A. Postural hygiene program to prevent low back pain. *Spine*. 2001;26:1280–6.
- Pellisé F, Puig O, Rivas A, Bagó J, Villanueva C. Low fusion rate after L5-S1 laparoscopic anterior lumbar interbody fusion using twin stand-alone carbon fiber cages. *Spine*. 2002;27:1665–9.
- Bagó J, Climent JM, Ey A, Pérez-Grueso FJ, Izquierdo E. The Spanish version of the SRS-22 patient questionnaire for idiopathic scoliosis: Transcultural adaptation and reliability analysis. *Spine*. 2004;29:1676–80.
- Martín-Benloch JA, Maruenda-Paulino JI, Barra-Pla A, Lagua-Garzarán M. Expansive laminoplasty as a method for managing cervical multilevel spondylotic myelopathy. *Spine*. 2003;28:680–4.
- Fernández-de-las-Peñas C, Downey C, Miangolarra-Page JC. Validity of the lateral gliding test as tool for the diagnosis of intervertebral joint dysfunction in the lower cervical spine. *J Manipulative Physiol Ther*. 2005;28:610–6.
- Pellisé F, Vidal X, Hernández A, Cedraschi C, Bagó J, Villanueva C. Reliability of retrospective clinical data to evaluate the effectiveness of lumbar fusion in chronic low back pain. *Spine*. 2005;30:365–8.
- Vicario C, López-Oliva F, Sánchez-Lorente T, Zimmermann M, Asenjo-Siguero JJ, Ladero F, et al. Arthrodesis cervical anterior mediante implante de tantalio. Resultados clínicos y radiológicos. *Neurocirugía (Astur)*. 2006;17:132–9.
- Álvarez E, Llargo A, Vaquero M, Lachica E. Modelo predictivo de la duración de incapacidad temporal por lumbalgia. Factores determinantes. *Revista de la Asociación Española de Especialistas en Medicina del Trabajo* 2005;14:106–13.
- Arana E, Martí-Bonmatí L, Vega M, Bautista D, Mollá E, Costa S, et al. Relationship between low back pain, disability, MR imaging findings and health care provider. *Skeletal Radiol*. 2006;35:641–7.
- Arana E, Martí-Bonmatí L, Montijano R, Bautista D, Molla E, Costa S. Relationship between Northwick Park neck pain questionnaire and cervical spine MR imaging findings. *Eur Spine J*. 2006;15:1183–8.
- Arana E, Martí-Bonmatí L, Mollá E, Costa S. Upper thoracic-spine disc degeneration in patients with cervical pain. *Skeletal Radiol*. 2004;33:29–33.
- Arana E, Martí-Bonmatí L, Dosdá R, Mollá E. Concomitant lower thoracic spine disc disease in lumbar spine MR imaging studies. *Eur Radiol*. 2002;12:2794–8.
- Barriga A, Díaz de Rada P, Barroso JL, Alfonso M, Lamata M, Hernáez S, et al. Frozen cancellous bone allografts: Positive cultures of implanted grafts in posterior fusions of the spine. *Eur Spine J*. 2004;13:152–6.
- Bagó J, Ramírez M, Pellisé F, Villanueva C. Survivorship analysis of Cotrel-Dubouset instrumentation in idiopathic scoliosis. *Eur Spine J*. 2003;12:435–9.
- Beguiristáin JL, Díaz-de-Rada P. Spondylolisthesis in pre-school children. *J Pediatr Orthop B*. 2004;13:225–30.
- Beguiristáin JL, Villas C, Preite R, Martínez R, Barrios RH. Lumbosacral arthrodesis using pedicular screws and ringed rods. *Eur Spine J*. 1997;6:233–8.
- Beguiristáin JL, Martínez Peric R, Barrios RH, Villas C. Lumbosacral arthrodesis with lous technique. Review of 186 cases. *Eur Spine J*. 1994;3:169–71.
- Colomina MJ, Godet C, Pellisé F, Bagó J, Villanueva C. Transcranial Doppler monitoring during laparoscopic anterior lumbar interbody fusion. *Anesth Analg*. 2003;97:1675–9.
- De la Torre-Gutiérrez M, Martínez-Quiñones JV, Escobar-Solís R, de la Torre-Gutiérrez S. Navegación raquídea. Nuestra experiencia. *Neurocirugía (Astur)*. 2001;12:490–8.
- Escala J, Huguet R, Giné J. Quimionucleolisis. Revisión a 10 años. *Rev Ortop Traumatol*. 1999;43:221–5.
- Faile I, González JL, Novallbos JP, Casais F, Marín J, Elorza J. Psychological and occupational predictive factors for back pain among employees of a university hospital in southern Spain. *Occup Med (Lond)*. 2000;50:591–6.
- Ferrández L, Usabiaga J, Ramos L, Arévalo MA, Domínguez J. Clinical and ultrastructural analysis of failures in chemonucleolysis. *Clin Orthop Relat Res*. 1992;283:74–80.
- Ferrer P, Martí-Bonmatí L, Mollá E, Arana E. MR-myelography as an adjunct to the MR examination of the degenerative spine. *MAGMA*. 2004;16:203–10.
- Ferrer P, Martí-Bonmatí L, Mollá E, Casillas C. Utilidad de la mielografía mediante resonancia magnética en el diagnóstico de las enfermedades de la columna vertebral. *Med Clin (Barc)*. 2000;115:366–9.
- Gómez Suárez E, Quintana Gómez JL, Romero De Lama C, De Felipe Medina R. Lumbalgia e incapacidad transitoria. *Revista de la Asociación Española de Especialistas en Medicina del Trabajo* 2005;14:8–11.
- González-Pedrouzo JE, Marsol Puig A, Huguet-Comelles R, Provincial-Fatsini E, Giné-Goma J. Factores pronósticos del tratamiento quirúrgico de la hernia discal lumbar. *Rev Ortop Traumatol*. 2004;41:351–6.
- González Viejo MA, Oyarzábal A, Fernández Fernández de Leceta Z, Rotinen J, Arrizabalaga N, Ibáñez de Garayo E. Satisfacción de los pacientes con cervicalgia y lumbalgia en un servicio de rehabilitación. *An Sist Sanit Navar*. 2005;28:379–88.
- González Viejo MA, Condón Huerta MJ. Incapacidad por dolor lumbar en España. *Med Clin (Barc)*. 2000;114:491–2.
- Gros Bañeres B, Bertol Alegre V, Fernández García A, Mallen Belenguer M, García Aranda C, Ginés García C. Las lumbalgias y su impacto en atención primaria. *Aten Primaria*. 1992;9:319–21.
- Gutiérrez Rubio A, Del Barrio Mendoza A, Ruiz Frutos C. Factores de riesgo y patología lumbar ocupacional. *Mapfre Medicina*. 2001;12:204–13.
- Hernández A, Pellisé F, Becerra JA, Rovira A, Bagó J, Villanueva C. Signos predictivos de dolor discogénico lumbar: correlación de la RM con la discografía. *Rev Ortop Traumatol*. 2002;46:528–33.
- Isla A, Álvarez F. Fibrosis epidural espinal post-discectomía lumbar y barrera anti-adhesiva. *Neurocirugía (Astur)*. 2001;12:439–46.
- Lladó A, Sologaitua E, Guimerá J, Marín M. Expanded polytetrafluoroethylene membrane for the prevention of peridural fibrosis after spinal surgery: A clinical study. *Eur Spine J*. 1999;8:144–50.
- Marín FZ. CAM versus nucleoplasty. *Acta Neurochir Suppl*. 2005;92:111–4.
- Martín R, Carda JR, Montaña F, Pinto JI, Sanz F, Paternina B, et al. Mielopatía cervical: análisis retrospectivo de los resultados quirúrgicos de 54 pacientes tratados mediante discectomía y fusión intersomática por vía anterior. *Neurocirugía (Astur)*. 2005;16:235–55.
- Martínez-Segura R, Fernández-de-las-Peñas C, Ruiz-Sáez M, López-Jiménez C, Rodríguez-Blanco C. Immediate effects on neck pain and active range of motion after a single cervical high-velocity low-amplitude manipulation in subjects presenting with mechanical neck pain: A randomized controlled trial. *J Manipulative Physiol Ther*. 2006;29:511–7.
- Mollá E, Martí-Bonmatí L, Arana E, Martínez-Bisbal MC, Costa S. Magnetic resonance myelography evaluation of the lumbar spine end plates and intervertebral disks. *Acta Radiol*. 2005;46:83–8.
- Pellisé F, Toribio J, Rivas A, García-Gontecha C, Bagó J, Villanueva C. Clinical and CT scan evaluation after direct defect repair in spondylolysis using segmental pedicular screw hook fixation. *J Spinal Disord*. 1999;12:363–7.
- Pujol Robinat A, Puig L, Mansilla J, Idiáquez I. Síndrome del latigazo cervical: factores relevantes en el pronóstico médico-legal. *Med Clin (Barc)*. 2003;121:109–15.
- Roca J, Iborra M, Cavanilles-Walker JM, Albertí G. Direct repair of spondylolysis using a new pedicle screw hook fixation: Clinical and CT-assessed study: An analysis of 19 patients. *J Spinal Disord Tech*. 2005;18:S82–9.
- Roca J, Ubierna MT, Cáceres E, Iborra M. One-stage decompression and posterolateral and interbody fusion for severe spondylolisthesis. An analysis of 14 patients. *Spine*. 1999;24:709–14.
- Royo-Salvador MB, Sabaté C, Monteiro A, Gil A, Ruiz R, Querol J, et al. Hernia discal lumbar en el ámbito laboral. Resultado de un análisis retrospectivo de una serie de 189 pacientes consecutivos. *Rev Neurol*. 1998;27:574–6.
- Ruiz Santiago FR, Milena GL, Herrera RO, Romero PA, Plazas PG. Morphometry of the lower lumbar vertebrae in patients with and without low back pain. *Eur Spine J*. 2001;10:228–33.
- Sanpera Jr I, Beguiristáin-Gurpide JL. Bone scan as a screening tool in children and adolescents with back pain. *J Pediatr Orthop*. 2006;26:221–5.
- Sabaté López P, Bestratén Bellobi J, Llor Vila JL, Santigosa Benet J, Murria Pérez MJ, Milá Niubó A, et al. Estudio de las lumbalgias atendidas en un centro de salud. *Aten Primaria*. 1992;9:208–11.
- Sarasqueta C, Gabaldon O, Iza I, Béland F, Paz PM. Cross-cultural adaptation and validation of the NASS outcomes instrument in Spanish patients with low back pain. *Eur Spine J*. 2005;14:586–94.
- Saune Castillo M, Arias Anglada R, Lletget Maymó I, Ruiz Bassols A, Escriba Jordana JM, Gil M. Estudio epidemiológico de la lumbalgia. Análisis de factores predictivos de incapacidad. *Rehabilitación (Madr)*. 2003;37:3–10.
- Pinto-Meza A, Serrano-Blanco A, Codony M, Reneses B, Von Korff M, Haro JM, et al. Prevalencia y comorbilidad física y mental del dolor dorsal y cervical crónicos en España: resultados del estudio ESEMeD. *Med Clin (Barc)*. 2006;127:325–30.
- Ferrer M, Pellisé F, Escudero O, Álvarez L, Pont A, Alonso J, et al. Validation of a minimum outcome core set in the evaluation of patients with back pain. *Spine*. 2006;31:1372–9.
- Muñoz M, García-Vallejo JJ, Sempere JM, Romero R, Olalla E, Sebastián C. Acute phase response in patients undergoing lumbar spinal surgery: Modulation by

- perioperative treatment with naproxen and famotidine. *Eur Spine J*. 2004;13:367–73.
55. Florez García M, García Pérez MA, García Pérez F, Armenteros Pedreros J, Álvarez Prado A, Martínez Lorente MD. Adaptación transcultural a la población española de la escala de incapacidad por dolor lumbar de Oswestry. *Rehabilitación*. 1995;29:138–54.
 56. Bassols A, Bosch F, Campillo M, Baños JE. El dolor de espalda en la población catalana. Prevalencia, características y conducta terapéutica. *Gac Sanit*. 2003;17:97–107.
 57. González-Urzelai V, Palacio-Elua L, López-de-Munain J. Routine primary care management of acute low back pain: Adherence to clinical guidelines. *Eur Spine J*. 2003;12:589–94.
 58. Gómez-León N, Rodríguez-González R, Calzado-Cantera A. Evaluación de la calidad de los exámenes de tomografía computarizada de columna lumbar en la indicación de hernia discal. *Radiología*. 2004;46:219–24.
 59. Vas J, Perea-Milla E, Méndez C, Silva LC, Herrera Galante A, Aranda Regules JM, et al. Efficacy and safety of acupuncture for the treatment of non-specific acute low back pain: a randomised controlled multicentre trial protocol [ISRCTN65814467]. *BMC Complement Altern Med*. 2006;6:14.
 60. Vas J, Perea-Milla E, Méndez C, Sánchez Navarro C, León Rubio JM, Brioso M, et al. Efficacy and safety of acupuncture for chronic uncomplicated neck pain: A randomised controlled study. *Pain*. 2006;126:245–55.
 61. Arnau JM, Vallano A, López A, Pellisé F, Delgado MJ, Prat N. A critical review of guidelines for low back pain treatment. *Eur Spine J*. 2006;15:543–53.
 62. Corcoll J, Orfila J, Tobajas P, Alegre L. Implementation of neuroreflexotherapy for subacute and chronic neck and back pain within the Spanish public health system: Audit results after one year. *Health Policy*. 2006;79:345–57.
 63. Moreno J, Gestoso M, Kovacs FM. La efectividad de la intervención neuroreflejojoterápica en el tratamiento de la lumbalgia inespecífica: resultados preliminares. *Medicina del Trabajo*. 1992;1:433–43.
 64. Kovacs FM, Abraira V, López-Abente G, Pozo F. La intervención neuroreflejojoterápica en el tratamiento de la lumbalgia inespecífica: un ensayo clínico controlado, aleatorizado, a doble ciego. *Med Clin (Barc)*. 1993;101:570–5.
 65. Kovacs FM, Abraira V, Pozo F, Kleinbaum DG, Beltrán J, Mateo I, et al. Local and remote sustained trigger point therapy for exacerbations of chronic low back pain. A randomized, double-blind, controlled, multicenter trial. *Spine*. 1997;22:786–97.
 66. Gil del Real MT, Kovacs FM, Gestoso M, Mufraggi N, Diéguez JM, the Balearic Back Pain Group. Evaluation of two questionnaires to determine exposure to risk factors for non-specific low back pain in Mallorcan schoolchildren and their parents. *Eur J Public Health*. 1999;9:194–9.
 67. Kovacs FM, Llobera J, Gil Del Real MT, Abraira V, Gestoso M, Fernández C, Primaria Group KA. Validation of the Spanish version of the Roland-Morris questionnaire. *Spine*. 2002;27:538–42.
 68. Kovacs FM, Llobera J, Abraira V, Lázaro P, Pozo F, Kleinbaum D, KAP Group. Effectiveness and cost-effectiveness analysis of neuroreflexotherapy for subacute and chronic low back pain in routine general practice: A cluster randomized, controlled trial. *Spine*. 2002;27:1149–59.
 69. Kovacs FM, Gestoso M, Gil del Real MT, López J, Mufraggi N, Méndez JI. Risk factors for non-specific low back pain in schoolchildren and their parents: A population based study. *Pain*. 2003;103:259–68.
 70. Kovacs FM, Abraira V, Peña A, Martín-Rodríguez JG, Sánchez-Vera M, Ferrer E, et al. Effect of firmness of mattress on chronic non-specific low-back pain: Randomised, double-blind, controlled, multicentre trial. *Lancet*. 2003;362:1599–604.
 71. Kovacs FM, Abraira V, Zamora J, Teresa Gil del Real M, Llobera J, Fernández C, Atención Primaria Group. Correlation between pain, disability, and quality of life in patients with common low back pain. *Spine*. 2004;29:206–10.
 72. Grupo PINS. Intervención neuroreflejojoterápica para el tratamiento de las enfermedades mecánicas del raquis. Resultados de una experiencia piloto. *Gac Sanit*. 2004;18:275–86.
 73. Kovacs FM, Abraira V, Zamora J, Fernández C, Spanish Back Pain Research Network. The transition from acute to subacute and chronic low back pain: A study based on determinants of quality of life and prediction of chronic disability. *Spine*. 2005;30:1786–92.
 74. Kovacs FM, Muriel A, Abraira V, Medina JM, Castillo Sánchez MD, Olabe J, Spanish Back Pain Research Network. The influence of fear avoidance beliefs on disability and quality of life is sparse in Spanish low back pain patients. *Spine*. 2005;30:E676–82.
 75. Kovacs FM, Muriel A, Medina JM, Abraira V, Sánchez MD, Jaúregui JO, Spanish Back Pain Research Network. Psychometric characteristics of the Spanish version of the FAB questionnaire. *Spine*. 2006;31:104–10.
 76. Kovacs FM, Fernández C, Cordero A, Muriel A, González-Luján L, Gil del Real MT, Spanish Back Pain Research Network. Non-specific low back pain in primary care in the Spanish National Health Service: A prospective study on clinical outcomes and determinants of management. *BMC Health Serv Res*. 2006;6:57.
 77. Zamora J, Kovacs F, Abraira V, Fernández C, Lázaro P, The Spanish Back Pain Research Network. The social tariff of EQ-5D is not adequate to assess quality of life in patients with low back pain. *Qual Life Res*. 2006;16:523–31.
 78. Kovacs FM, Urrútia G, Nishinshiya B, Olabe J, Roque M, García J, et al. Systematic review on the efficacy and safety of intervertebral disc prostheses in the lumbar spine. *Spine J*. 2006;6:4S–5S.
 79. Aalto TJ, Malmivaara A, Kovacs F, Herno A, Alen M, Salmi L, et al. Preoperative predictors for postoperative clinical outcome in lumbar spinal stenosis: Systematic review. *Spine*. 2006;31:E648–63.
 80. Urrútia G, Burton AK, Morral A, Bonfill X, Zanoli G. Neuroreflexoterapia para el tratamiento del dolor lumbar inespecífico (Revisión Cochrane traducida). En: *La Biblioteca Cochrane Plus*, 2008 Número 4. Oxford: Update Software Ltd.
 81. Van Tulder M, Becker A, Bekkering A, Breen A, Gil del Real MT, Hutchinsonson A, et al. European guidelines for the management of acute nonspecific low back pain in primary care. *Eur Spine J*. 2006;15:S169–91.
 82. Airkasininen O, Brox JJ, Cedraschi C, Hildebrandt J, Klüber-Moffet J, Kovacs FM, et al. European guidelines for the management of chronic nonspecific low back pain. *Eur Spine J*. 2006;15:S192–300.
 83. Pérez Irazusta I, Alcorta Michelena I, Aguirre Lejarcegui G, Aristegi Racero G, Caso Martínez J, Esquisabel Martínez R, et al. Guía de práctica clínica sobre lumbalgia, Osakidetza. GPC 2007/1. Vitoria-Gasteiz [consultado 26/03/2009]. Disponible en: <http://www.guiasalud.es>.
 84. Kovacs FM, Muriel A, Castillo Sánchez MD, Medina MJ, Royuela A, Spanish Back Pain Research Network. Fear avoidance beliefs influence duration of sick leave in Spanish low back pain patients. *Spine*. 2007;32:1761–6.
 85. Kovacs FM, Muriel A, Castillo Sánchez MD, Medina JM, Royuela A, Gil del Real MT, et al. Fear avoidance beliefs do not influence disability and quality of life in Spanish elderly subjects with low back pain. *Spine*. 2007;32:2133–8.
 86. Kovacs FM, Abraira V, Santos S, Díaz E, Gestoso M, Muriel A, et al. A comparison of two short education programs for improving low back pain-related disability in the elderly. A cluster randomized controlled trial. *Spine*. 2007;32:1053–9.
 87. Kovacs FM, Bagó J, Royuela A, Seco J, Giménez S, Spanish Back Pain Research Network. Psychometric characteristics of the Spanish version of instruments to measure neck pain disability. *BMC Musculoskelet Disord*. 2008;9:42. [consultado 20/03/2009]. Disponible en: www.kovacs.org.
 89. Carragee EJ, Deyo RA, Kovacs FM, Peul WC, Lurie JD, Urrútia G, et al. Clinical research: Is the spine field a mine field? *Spine*. 2009;5:423–30.

6. Discussió



6.- DISCUSSIÓ

6.1.- Breu discussió específica derivada de les publicacions

Neurorreflexoteràpia per la lumbàlgia no específica

La revisió dels assaigs clínics publicats mostra que la neurorreflexoteràpia (NRT) és una intervenció eficaç (millor que placebo i que el maneig habitual a primària) per al tractament de la lumbàlgia crònica inespecífica. Possiblement ho sigui també per la lumbàlgia sub-aguda, però no hi ha prou dades per confirmar-ho o descartar-ho.

Dels estudis no és possible saber quant temps es manté el seu efecte ni el percentatge de recurrències a mig i llarg termini, encara que un seguiment addicional a un any només dels pacients del grup d'intervenció en un dels AC (dades publicades però no incloses a la revisió) suggereix que l'efecte de la NRT es manté a l'any.

La qualitat dels estudis, la consistència dels resultats i la magnitud de l'efecte observat permeten afirmar que la NRT té un efecte favorable sobre el dolor, la discapacitat i la funció, d'inici molt precoç (immediatament després de la seva aplicació) i que es manté al menys fins als 60 dies.

No sembla que la NRT tingui problemes de seguretat associats. Malgrat el que pugui semblar, es tracta d'un procediment 'invasiu' poc agressiu que provoca poques molèsties als pacients. A més, ha demostrat ser cost-efectiva respecte a l'atenció habitual pel metge de capçalera, en termes d'una reducció de costos sanitaris (visites, proves i tractaments). Cal tenir en compte que l'atenció habitual en aquest context específic no implica la utilització d'un protocol de maneig assistencial més o menys estandarditzat a l'atenció primària (de fet, es tracta d'una patologia on hi ha molta variabilitat), sinó que reproduïx la situació de la pràctica clínica real a l'atenció primària on les decisions sobre el maneig

d'aquesta patologia es prenen segons el millor criteri de cada metge i on s'observa una elevada variabilitat.

Com la majoria d'intervencions no farmacològiques, el seu efecte és també terapeuta-depenent ja que requereix un alt grau d'ensinistrament, especialment pel que fa a la correcta identificació del/s nivell/s anatòmic/s implicats en la gènesi del dolor i la correcta aplicació de les grapes en el lloc adient.

La magnitud del benefici observat amb la NRT respecte al placebo és inusualment elevat en comparació amb d'altres tractaments disponibles per la lumbàlgia. Això, que genera suspicàcies en alguns, indicaria que la NRT és una intervenció realment molt efectiva però seria aconsellable comparar-la amb altres opcions d'eficàcia ben demostrada, sobretot en el cas de voler aplicar-la de forma generalitzada a l'àmbit del sistema nacional de salut. No obstant, la principal responsabilitat de promoure aquests estudis sobre efectivitat clínica recau, fonamentalment, en el sistema que hauria de facilitar la recerca que permeti identificar les millors opcions terapèutiques, especialment per a problemes de salut molt rellevants i/o amb un elevat impacte.

Fins a la data, la NRT només és aplicada per un grup reduït de terapeutes acreditats a l'estat espanyol, alguns d'ells en el sistema nacional de salut a diverses comunitats autònomes. Es fa difícil, doncs, pensar en la possibilitat de d'una difusió a d'altres àmbits geogràfics. Per aquest motiu, les guies de pràctica clínica internacionals sobre lumbàlgia només la recomanen allà on està disponible.

Una actualització encara no publicada d'aquesta revisió va identificar una publicació posterior consistent en un estudi d'avaluació de resultats (*audit*) a llarg termini (un any) per part del Servei de Salut de les Illes Balears.⁽⁶²⁾ L'estudi es va portar a terme per recolzar la decisió de continuar o no amb la prestació amb finançament públic de la NRT al sistema de salut balear. L'estudi avalua els resultats clínics a un any dels primers 1.209 pacients consecutius (casos no seleccionats) remesos pels seus metges de capçalera a una de les tres unitats de NRT de referència de les Illes. Aquest protocol d'avaluació havia estat prèviament consensuat com a condició per a la implantació de la NRT a les Balears. A més, s'avalua l'adequació en l'aplicació dels criteris de remissió

establerts al protocol per part dels metges i la satisfacció de professionals i pacients. El percentatge de pacients correctament remesos va ser del 95,5%, la qual cosa suggereix que es tracta d'un protocol senzill i factible d'aplicar en el context de les consultes de primària. La NRT només va ser rebutjada pel 2,7% dels pacients. La reducció significativa del dolor d'esquena i referit a la cama (mesurada amb VAS) i de la discapacitat (mesurada amb el qüestionari Roland Morris) observades en els AC a curt termini es varen mantenir a l'any, sense que es comunicessin efectes adversos. La satisfacció de metges i pacients va ser elevada. Això situa la NRT com un tractament prometedor que convindria tenir en compte com a opció terapèutica, especialment en les àrees on existeixen professionals degudament acreditats. A més, la NRT es tracta d'una tecnologia poc sofisticada (empra grapes quirúrgiques convencionals) i, per tant, poc costosa. Actualment s'està portant a terme un estudi pilot a Catalunya.

IDET per la lumbàlgia discogènica

Els resultats conflictius dels estudis de la revisió no recolzen l'ús de les tècniques de termocoagulació intradiscal percutànies per al tractament de la lumbàlgia d'origen discogènic.

La revisió posa de manifest com la qualitat dels estudis originals afecta els seus resultats, en el sentit de que els estudis de baixa qualitat i alt risc de biaix tendeixen a sobreestimar l'efecte dels tractaments. Això és particularment cert quan els qui apliquen el tractament tenen un clar conflicte d'interès (econòmic i/o professional), com és el cas de la IDET i de moltes altres intervencions per al tractament de la lumbàlgia. Per això, la selecció no esbiaixada de pacients (ocultació de la seqüència de les assignacions), l'emascament de les intervencions i/o l'avaluació cegada dels resultats resulten aspectes fonamentals per garantir la fiabilitat d'aquests tipus d'estudi.

Els estudis de baixa qualitat mostraven resultats favorables, com també els havien observat abans les nombroses sèries de casos publicades a la literatura, moltes d'elles per part de professionals de l'àmbit de la medicina privada, especialment dels EEUU. Hi ha moltes raons per sospitar que aquests

resultats estan esbiaixats. Per contra, els estudis de millor qualitat de la revisió no demostren que l'IDET sigui eficaç comparada amb placebo (IDET simulada). Fins i tot en un altre AC s'observa com l'efecte inicial d'aquesta intervenció és fugaç.

Els pacients amb lumbàlgia d'origen discogènic varen ser seleccionats en base a uns criteris diagnòstics qüestionables, concretament pel que fa a l'ús de la discografia de provocació o analgèsica. Aquesta prova ha estat qüestionada perquè dona molts resultats falsos positius, la qual cosa provoca una mala classificació i selecció de pacients. Encara que s'ha proposat emprar la ressonància magnètica per millorar la precisió diagnòstica de la discografia, tampoc hi ha consens de que aquesta sigui una prova diagnòstica òptima ja que no sempre hi ha una bona correlació entre els resultats de la prova i els símptomes. En conseqüència, la lumbàlgia d'origen discogènic consistiria en una entitat de difícil categorització que ha portat a alguns a qüestionar la seva veritable entitat. Això implica, també, la dificultat per dur a terme estudis de qualitat mentre no es disposi de mètodes de diagnòstic adequats que identifiquin correctament l'origen discogènic del dolor.

Un altre fet rellevant detectat per la revisió són els nombrosos problemes metodològics que presenten els assaigs clínics d'intervencions no farmacològiques, que limiten la possibilitat de concloure sobre la seva veritable eficàcia amb un alt grau de certesa (aquestes limitacions es comenten més endavant a la part de discussió general). Això no suposa pas una limitació de la revisió com a mètode de recerca i avaluació de les intervencions, sinó més aviat al contrari, que la revisió es veu limitada per la qualitat de les dades disponibles. Una bona revisió sistemàtica no resol els problemes d'estudis deficients, simplement identifica el problema i permet ajustar i modular la interpretació dels resultats tenint en compte la base d'evidència de que es disposa.

Finalment, en el cas de la IDET no es pot descartar del tot que no tingui cap efecte favorable. Simplement, els estudis disponibles no han estat capaços de demostrar-ho. Tenint en compte que es tracta d'un procediment mínimament invasiu, seria recomanable continuar investigant, superant però, les nombroses

mancances metodològiques ja apuntades (correcte diagnòstic, mostra àmplia de pacients, comparació primer amb placebo [simulació] per demostrar eficàcia i, en cas positiu, amb un control adient que hagi demostrat prèviament la seva eficàcia, seguiment llarg, registre acurat dels efectes adversos i, idealment, amb participació de diversos centres i investigadors sense conflictes d'interès).

Cirurgia per l'estenosi espinal lumbar simptomàtica

La revisió suggereix que la cirurgia de descompressió és efectiva en termes de reducció del dolor i millora de la funció i de la qualitat de vida, en comparació amb el tractament conservador, en pacients amb dolor radicular produït per estenosi espinal i que han fracassat a tractament previ per 3 a 6 mesos.

La cirurgia sembla ser superior al tractament conservador independentment de la presència o no d'espondilolistesi o de claudicació neurogènica, encara que el benefici observat sembla major en presència de claudicació.

L'efecte positiu de la cirurgia sembla persistir encara als 4 anys i, possiblement, també als 10 anys, tal com apunten alguns estudis de seguiment de cohort prospectiva.⁽⁶³⁾ De confirmar-se, es tractaria d'un benefici que justificaria el risc quirúrgic que comporta la intervenció.

L'addició de fusió a la descompressió estaria justificada només en presència d'espondilolistesi, sense diferències entre el tipus de fusió.

La indicació de la cirurgia implica una acurada valoració del risc individual del pacient i de la seva expectativa de benefici.

Els resultats dels pacients operats a l'inici i aquells que varen creuar-se del grup control a cirurgia en el decurs de l'estudi varen ser similars, indicant que el retard en la cirurgia no s'associa amb un pitjor pronòstic.

L'implant interespinal de tracció X-STOP va mostrar uns resultats igualment favorables i és força menys agressiu que la cirurgia de descompressió (es pot fer de manera ambulatoria), amb menor risc de sagnat i també risc operatori. No obstant això, es tracta d'un únic estudi i amb conflictes d'interès, per la qual

cosa caldria disposar de més dades, idealment amb més estudis ben dissenyats, per confirmar la seva eficàcia i seguretat també a llarg termini. Donats els conflictes d'interès implicats en aquesta tecnologia, els estudis haurien de poder garantir una avaluació imparcial dels resultats.

Atesa la naturalesa de la pregunta de recerca i el seu context, tots els estudis de la revisió varen seguir un plantejament pragmàtic a l'hora de decidir el tipus de cirurgia i el tractament conservador del grup control. Aquests es varen deixar a la lliure elecció del metge responsable davant de cada cas, sense que s'establís un estàndard comú per a tots els pacients. Aquesta heterogeneïtat clínica reflecteix la variabilitat que existeix a la pràctica clínica real. Això que podria haver suposat un problema important per a la interpretació dels resultats, no ho va ser tant donada la consistència dels resultats entre els estudis.

La revisió mostra alguns problemes habituals en els assaigs clínics que avaluen la cirurgia i que poden influir en alguna mesura en els resultats. El rebuig dels pacients a participar en un assaig clínic on se'ls proposa l'opció de rebre tractament conservador, sobretot quan ja l'han provat sense èxit i tenen unes altes expectatives sobre la cirurgia, pot afectar la composició de la mostra final d'estudi (representativitat de la mostra) així com augmentar la taxa d'abandonaments i de creuaments en el decurs de l'estudi (la taxa de creuaments va ser molt alta en alguns estudis, de fins al 54%). Això planteja reptes importants a la fase d'anàlisi (anàlisi per ITT versus anàlisi segons tractament rebut) i interpretació dels resultats, com va mostrar algun dels estudis més importants de la revisió. En aquest cas particular, es va estimar convenient considerar també l'anàlisi segons el tractament rebut si es volia veure l'efecte específic del tractament quirúrgic, ja que l'anàlisi per ITT avalua l'efecte d'una 'política' (la decisió de donar un tractament enlloc d'un altre) més que no pas d'un tractament. Com es va posar de manifest als estudis SPORT inclosos en la revisió^(64,65), és evident que la condició clínica dels pacients pot canviar al llarg de l'estudi (especialment si aquest té un seguiment llarg i la malaltia és evolutiva, com és el cas de l'estenosi espinal), i això pot afectar l'adherència al tractament assignat fomentant el creuament entre grups d'estudi. El plantejament que fa l'anàlisi per ITT pot resultar inadequat o

insuficient en algunes determinades circumstàncies, conduint cap a una infraestimació o dilució de l'efecte de la intervenció, com pensem ha estat en aquest cas. El veritable efecte de la cirurgia es trobaria, doncs, en un punt mig entre la visió conservadora de l'anàlisi per ITT i la visió optimista de l'anàlisi segons el tractament rebut. No obstant això, els resultats obliguen a una reflexió i assenyalen la necessitat de dur a terme més investigació sobre el perfil específic dels pacients que més es poden beneficiar de la intervenció (factors pronòstics), per tal d'afinar els criteris d'indicació de la cirurgia.

La impossibilitat d'emascarar les intervencions i la manca de cegament en l'avaluació de resultats (la majoria de les variables eren de naturalesa subjectiva auto-reportats pels pacients) podrien haver introduït algun grau de biaix a favor de la cirurgia, en la mesura que existeixi la creença generalitzada entre els pacients de que la cirurgia és millor que el tractament conservador. No obstant, el seguiment ulterior d'aquests pacients a llarg termini proporciona evidència addicional de que l'efecte es manté en el temps, fet que difícilment podria atribuir-se a un efecte placebo.

En resum, l'evidència disponible sobre aquesta qüestió està lluny de ser la ideal. No obstant, tenint en compte els condicionants que dificulten la realització d'un assaig clínic en aquest context específic (cirurgia electiva versus tractament conservador), es pot concloure que és la millor possible en aquest camp. Per això, s'ha estimat suficient per concloure favorablement sobre l'efectivitat de la cirurgia, contràriament a la conclusió a què va arribar una revisió Cochrane anterior sobre aquest mateix tema.⁽⁶⁶⁾

No obstant això, romanen encara moltes qüestions obertes que justifiquen la necessitat d'investigar més en aquesta àrea, especialment en dues direccions: 1) cal investigar per identificar les característiques clíniques dels pacients que més influeixen en uns resultats favorables de la cirurgia per tal d'ajustar més les seves indicacions i millorar els seus resultats, i 2) cal dur a terme assaigs clínics que comparin les diferents tècniques quirúrgiques entre si (tipus de cirurgia, amb o sense fusió, tipus de fusió, etc.), i també comparant la cirurgia amb l'implant interespinós (en el cas de que els estudis actualment en curs confirmin els resultats favorables d'aquest dispositiu).

Ultrasons i ones de xoc per la lumbàlgia

Els resultats de la revisió no recolzen l'ús dels ultrasons i les ones de xoc en el tractament de la lumbàlgia.

Aquesta revisió posa de manifest un problema important i freqüent en l'avaluació de nombroses intervencions com és la selecció inadequada del tractament de comparació o control. Aquest hauria de ser un tractament que prèviament hagi demostrat de forma clara la seva eficàcia mitjançant assaigs clínics comparats amb placebo i que, a més, estigui disponible a la pràctica habitual. No obstant, un dels estudis de la revisió va comparar els ultrasons amb làser i tracció, sense que cap d'ells hagi demostrat la seva eficàcia en el tractament de la lumbàlgia. Conseqüentment, els resultats que suggereixen que els ultrasons són equivalents a làser i a tracció en cap cas es poden interpretar com un recolzament de l'efectivitat de cap d'aquestes intervencions.

La majoria dels efectes detectats als estudis de la revisió varen ser molt petits i clínicament irrellevants. Aquest fet posa de manifest que no tot efecte positiu, en el sentit d'una diferència estadísticament significativa, és clínicament rellevant i podria justificar un canvi d'actitud terapèutica. La manca d'informació als estudis sobre la diferència mínima clínicament rellevant, ja sigui expressada de forma directa o indirecta a partir de la justificació del càlcul de la mida mostral, pot dificultar aquesta valoració sobre la importància dels resultats. Això és particularment problemàtic quan no hi ha un consens general al respecte. No obstant, quan es comparen tractaments actius entre sí, resulta particularment important (i difícil!) pre-especificar la diferència mínima que cal observar per jutjar la superioritat d'un tractament sobre un altre. Així com per determinar eficàcia ('el tractament té un efecte positiu'), aquesta diferència pot tenir un valor fix (per exemple, la reducció en la puntuació del VAS per decidir que hi ha un efecte analgèsic), per determinar la superioritat d'un tractament sobre un altre aquesta diferència és relativa, ja que depèn també d'altres aspectes del tractament, com la seguretat, comoditat, accessibilitat o costos, molt variables entre tractaments.

Novament, els estudis d'aquesta revisió varen presentar nombroses limitacions metodològiques que es comenten de manera general més endavant.

Val la pena comentar que aquesta revisió recolza les principals guies de pràctica clínica basades en l'evidència sobre lumbàlgia que no recomanen l'ús d'aquest procediment.

Investigació espanyola sobre les patologies de l'esquena

L'estudi de la investigació espanyola sobre les patologies de l'esquena presenta importants limitacions (especialment, per la baixa taxa de resposta als qüestionaris enviats per correu a les entitats, centres i investigadors) que poden posar en dubte l'exactitud de les dades aportades però no així les conclusions generals. No era la intenció de l'estudi fer un registre exhaustiu de tots els estudis i les seves publicacions. Tampoc les dades del finançament segurament són del tot exactes. No obstant, el missatge central de l'estudi no es troba contingut en les xifres sinó en les tendències que aquestes mostren.

En primer lloc, resulta difícil determinar si el nombre de projectes sobre patologies de l'esquena finançats per entitats espanyoles és alt o baix. Per això, caldria disposar del denominador (tots els projectes de recerca finançats per aquestes entitats). No obstant, tenint en compte l'elevat impacte social i sanitari d'aquestes patologies, arribem a la conclusió raonable de que molt probablement sigui baix.

Aquesta conclusió es reforça si tenim en compte que un elevat percentatge d'aquests projectes han estat finançats per una mateixa entitat privada sense ànim de lucre, com és la Fundación Kovacs. Això no ha de resultar gens estrany si pensem que la Fundación Kovacs està especialitzada en les patologies de l'esquena i coordina una àmplia xarxa d'àmbit estatal i caràcter multidisciplinari d'investigadors en aquesta àrea coneguda amb el nom de REIDE (Red Española de Investigación en Dolencias de la Espalda), una xarxa altament productiva com es pot comprovar al seu web.⁽⁶⁷⁾ Per contrast, el que sí resulta cridaner és el baix nombre de projectes que han merescut ser finançats pels organismes públics de recerca, entre els quals es troben alguns dels

principals al nostre país. Encara que la xifra real sigui més alta, segurament en resulta encara un nombre massa baix per la importància d'aquesta patologia.

No ha estat possible conèixer el nombre total dels projectes presentats i denegats per aquestes entitats. Potser hagin estat pocs. En aquest hipotètic cas, caldria pensar en la necessitat de desenvolupar estratègies que incentivin i promoguin la recerca en aquest àmbit per part dels nombrosos grups existents al nostre país, molts d'ells solvents.

L'estudi també posa de manifest la imperfecció en els criteris i mecanismes emprats per les entitats públiques finançadores de la recerca, que es demostra en dos aspectes diferents. D'una banda, l'elevat nombre de projectes finançats que després no es publiquen (recerca improductiva). La selecció d'un projecte en una convocatòria competitiva se suposa que és un indicador de la seva qualitat, rellevància i pertinença, així com també de la solvència de l'equip investigador. No obstant, la no publicació dels resultats semblaria indicar tot el contrari i, per tant, que els criteris de selecció no han estat del tot encertats. D'altra banda, i encara més sorprenent, molts dels estudis denegats per les entitats públiques, finalment varen poder ser realitzats amb l'ajuda d'una altra entitat privada i publicats, amb un factor d'impacte promig més elevat que els seleccionats. Això podria indicar que es tracta de projectes de major qualitat i/o interès que els seleccionats. Independentment de quines siguin aquestes entitats, aquesta dada posaria de manifest els criteris qüestionables aplicats en el procés de selecció dels projectes per part de les agències públiques. L'estudi no aprofundeix en les possibles causes, però vol ser un toc d'alerta perquè es revisin els criteris i mecanismes de selecció dels projectes finançats amb fons públics per tal de garantir, d'una banda, que és té prou en compte la rellevància dels problemes de salut a l'hora de prioritzar les àrees de recerca, i d'altra banda, la qualitat intrínseca dels projectes.

El reforçament d'aquests mecanismes i criteris és també important i necessari si es vol evitar l'elevat percentatge d'inversió improductiva que les xifres apunten. El sistema públic hauria de ser encara més transparent, no només en els seus criteris i mecanismes de selecció, sinó també en la informació sobre la gestió dels recursos (finançament de projectes) i els seus resultats (producció

científica). Ara per ara, la informació al respecte disponible al públic és escassa i parcial i, per tant, caldria avançar cap a una major transparència.

Tot i que l'estudi analitza dades d'un període passat (1992-2006), la seva amplitud permet pensar que es tracta d'una visió ponderada de l'estat de la qüestió a l'estat espanyol i res fa pensar que la situació hagi canviat substancialment en els anys més recents. La manca de transparència abans indicada ens impedeix comprovar-ho.

En resum, les agència de finançament de la recerca haurien d'incrementar la seva inversió en aquesta àrea i també revisar i millorar els seus mecanismes d'avaluació dels projectes.

6.2.- Discussió dels aspectes generals

Utilitat, limitacions i reptes per dur a terme l'avaluació de les intervencions mitjançant revisions sistemàtiques

Les quatre revisions sistemàtiques presentades anteriorment mostren la utilitat de les revisions en la gestió del coneixement, en la mesura que faciliten la informació als qui la necessiten per poder prendre decisions ben informades.

La revisió sobre NRT va ser la primera revisió sistemàtica sobre aquesta intervenció. La seva metodologia rigorosa així com el fet de que es publicés en una font bibliogràfica àmpliament difosa com és la Biblioteca Cochrane va facilitar-ne la seva difusió i, per tant, que fos ràpidament coneguda i emprada pels responsables del Servei de Salut de les Illes Balears a l'hora de prendre la decisió sobre la continuació del finançament públic d'aquesta prestació a les Illes Balears (on s'havia començat a aplicar l'any 2002). Els decisors necessiten informació de qualitat sobre els efectes de les intervencions que, posada en context (especialment considerant els recursos disponibles, valors i preferències locals), hauria de facilitar la presa de decisions sanitàries que

afavoreixin l'obtenció d'uns resultats favorables en termes de salut i de forma cost-efectiva.

La revisió també va ser incorporada a les dues guies de pràctica clínica basades en l'evidència més recents sobre lumbàlgia ja citades anteriorment.^(1,17) Les dues guies qualifiquen la revisió com de bona qualitat i conclouen que la NRT és una opció terapèutica que es pot considerar en aquells àmbits on la NRT estigui ja disponible (no és el cas del Regne Unit). De fet, la guia COST B13 va ser adaptada per al context espanyol per un grup de treball multidisciplinari i també recomana la NRT en els casos en que estigui indicada (la versió *online* de la guia es pot consultar al web <http://www.kovacs.org/Imagenes/Guia%20Lumbalgia.pdf>). El nivell d'evidència que recolza l'afirmació de que la NRT es eficaç i segura per al tractament de la lumbàlgia va ser qualificat de 'Nivell A' i l'afirmació de que la NRT és efectiva i cost-efectiva per al tractament de la lumbàlgia amb un 'Nivell C'. El 'Nivell A' es va definir com el màxim nivell d'evidència que correspon a una afirmació basada en el resultat concordant de diversos estudis d'alta qualitat metodològica (o una revisió sistemàtica dels mateixos, que també és d'alta qualitat metodològica). El 'Nivell C' es va definir com una afirmació basada en el resultat d'un sol assaig clínic (d'alta o baixa qualitat), o resultats inconsistents en una revisió sistemàtica de múltiples assaigs clínics. En ambdós casos, l'evidència que sustentava aquestes afirmacions de la guia va ser la nostra revisió sistemàtica sobre NRT.

En canvi, la revisió sobre procediments de termocoagulació intradiscal percutanis (IDET i PIRFT) per al dolor lumbar d'origen discogènic va ser publicada l'any 2007, després de la publicació de la guia COST B13. La guia no recomana aquesta intervenció, i conclou que l'evidència sobre la seva eficàcia (versus placebo) és contradictòria ('Nivell C'). Aquesta afirmació, plenament concordant amb la nostra revisió, es basa en nombrosos estudis identificats, sense que hi hagués cap revisió sistemàtica sobre el tema publicada en el moment de fer la guia, la qual cosa sens dubte va suposar un treball addicional de síntesi per als autors de la guia. Eventualment, la revisió hauria pogut suposar un important estalvi de temps per arribar a la conclusió final.

Tampoc la guia de NICE publicada al 2009, però amb una cerca bibliogràfica que abastà fins a gener de 2006, inclou la nostra revisió i, en canvi, sí una altra publicada anteriorment.⁽⁶⁸⁾ Aquesta revisió només inclou estudis aleatoris i exclou els de disseny controlat però no aleatori, i que hagin comparat IDET o PIRFT versus placebo. Conseqüentment, només inclou tres dels sis estudis de la nostra revisió. No obstant, les conclusions de la guia ('No es recomana l'ús d'IDET o PIRFT'), basant-se en que els resultats dels estudis identificats per la revisió són molt contradictoris, són plenament concordants amb la nostra revisió.

Les altres dues revisions han estat publicades molt recentment (any 2011) i, per tant, és poc probable que hagin tingut l'impacte que s'espera que assolixin. No obstant, és de ressaltar que la revisió de cirurgia publicada el 15 de setembre de 2011 ha estat seleccionada com article de la setmana per ser comentat en el blog de la revista Spine pel seu editor⁽⁶⁹⁾, per la seva qualitat i interès. Sobre la revisió, l'editor diu:

“Many recent studies have demonstrated the benefit of surgery compared to non-operative treatment for spinal stenosis (SpS). In the September 15 issue, Dr. Kovacs and his colleagues from Spain published an updated Cochrane-style review comparing surgical and non-operative treatment for SpS. The last Cochrane review on the topic was published in 2005 and concluded that the evidence in support of surgical treatment for SpS remained inconclusive. Based on the more recent studies, the conclusion has now been changed to state that there is strong evidence demonstrating the superiority of surgical treatment. The study rested on four randomized controlled trials (RCTs) comparing decompression (plus or minus fusion in some studies) with non-operative treatment and one RCT comparing an interspinous device with non-operative treatment. In all studies, the surgery group did significantly better than the non-operative group, so the conclusion for the systematic review was clear. The authors pointed out the limitations of the studies—blinding was generally impossible, crossover occurred at a high rate in three of the studies, and surgical and non-operative treatment was variable—but, rather than condemning them for these inherent limitations, came to the common sense conclusion that these trials offered

“the best possible evidence in this field.” It is refreshing to read systematic reviews that acknowledge that the randomized, double-blinded, placebo-controlled trial is not possible when comparing elective surgery to non-operative treatment.

So is this article just rehashing yesterday’s news? In some ways, systematic reviews and meta-analyses are by definition doing just that. However, in this case, it is a useful rehashing as this article should effectively close the book on the question about the effectiveness of decompressive surgery for spinal stenosis. After years of academic arguing about a question to which most people in the spine field felt they knew the answer based on their clinical experience, we can move onto answering more specific questions whose answers might be more helpful to our individual patients. While we know that decompressive surgery works for the “average” SpS patient, we now need to determine the best treatment for the individual patients we see in the office, none of whom are the “average” patient studied by these RCTs. We all know that some patients do just fine without surgery, but it can be quite difficult to predict who they are going to be. The future of this field lies in using evidence-based medicine to individualize treatment decisions that lead to the best result for the patient with the least invasive form of treatment that is in line with their preferences. To do this will require large studies so that the effect of specific individual characteristics on outcomes can be determined. When we get to the point that outcomes from specific surgical and non-operative treatments can be predicted based on individual patient characteristics, patients will truly be experiencing “evidence-based medicine.” Dr. Kovacs and his colleagues should be congratulated on writing the definitive review on this topic and encouraging us to move forward to the next stage of research in this field.”

En conclusió, la revisió ha estat percebuda com a molt útil per part dels clínics i decisors en aquesta àrea, tant per les seves implicacions per a la pràctica clínica com per a la recerca.

Un altre aspecte a esmentar aquí són les limitacions i reptes a l'hora de dur a terme una revisió sistemàtica. La intenció de l'autor de la revisió sempre és

proporcionar una peça d'informació que sigui la manera més ràpida, més completa i més fiable d'estar ben informat sobre un determinat tema. Idealment, s'esperaria que la revisió clarifiqui de forma raonable totes les incerteses plantejades a la pregunta d'investigació. No obstant, aquesta propietat de la revisió estarà, en gran mesura, condicionada per la quantitat i qualitat de la informació disponible. Si no hi ha estudis, o aquests no són de prou qualitat, és gairebé impossible que la revisió pugui proporcionar una resposta concloent. En aquest cas, la revisió sol concloure que *“no hi ha evidència que demostrï...”*. Tot i que això no resol la qüestió de l'eficàcia/efectivitat d'una intervenció, és un bon criteri per decidir (de moment) no promoure l'ús d'un tractament que encara no ha demostrat suficientment el seu benefici. La manca d'evidència no s'ha de confondre amb l'evidència de manca d'efecte.

Algunes de les revisions d'aquesta tesi han pogut arribar a conclusions prou clares, tot i que sempre resten aspectes per clarificar. Per exemple, tant la revisió sobre NRT com la de cirurgia per l'estenosi espinal lumbar simptomàtica són prou clares en el sentit de que recolzen l'afirmació de que ambdues intervencions són efectives. En aquests casos, la quantitat i la qualitat de l'evidència disponible s'ha considerat suficient per arribar a aquesta conclusió. Això no vol dir que els estudis detectats siguin perfectes, però la valoració conjunta o integrada del nombre d'estudis disponibles, la seva qualitat i els seus resultats, han permès als autors arribar a aquesta conclusió.

D'altra banda, pel que fa a la revisió sobre ultrasons, tot i les mancances importants dels estudis detectats, la conclusió és que es tracta d'una intervenció que no és eficaç i, per tant, no s'hauria de recomanar. Tot i les importants mancances dels estudis, s'ha jutjat com a poc probable que aquest tractament pugui tenir algun efecte favorable a la lumbàlgia.

Finalment, a la revisió sobre les tècniques percutànies de termocoagulació intradiscal, la conclusió final és que l'evidència és contradictòria ('no es pot saber si és eficaç o no'). Aquesta conclusió, en gran part, ve condicionada per la qualitat dels estudis disponibles que és força limitada.

En general, es podria afirmar que la qualitat d'una majoria d'estudis analitzats a les quatre revisions han presentat importants problemes metodològics que, o

bé han impedit arribar a conclusions més clares, o bé han limitat la possibilitat d'aprofundir en la identificació d'aquells aspectes que més poden influir en els resultats: identificació de subgrups específics de pacients que es poden beneficiar del tractament, modalitats específiques d'aplicació del tractament, etc. Aquesta limitació, probablement, és comuna a gran part de les intervencions per al tractament de la lumbàlgia.

Una altra potencial limitació a l'hora de fer una revisió i interpretar els seus resultats és l'heterogeneïtat clínica observada entre els estudis. Per exemple, tractant-se d'una mateixa patologia, sovint hi havia diferències en els criteris d'inclusió entre els estudis que dificultaven la seva comparació. També les variacions en les modalitats d'aplicació d'una mateixa intervenció, o en la manera de mesurar determinats resultats, ho dificulten. Tot això ha estat observat a les nostres revisions.

Aquesta variabilitat clínica, juntament amb la manera molt desigual de presentar els resultats entre els estudis, han impedit dur a terme cap metanàlisi. La combinació estadística dels resultats a través dels estudis proporciona un resum quantitatiu dels resultats que facilita la seva interpretació per part del lector, especialment quan el nombre d'estudis és elevat i, per tant, el volum de resultats també. No obstant, molt sovint, com ha estat el nostre cas, això no és possible, de manera que els resultats s'han de presentar de forma narrativa.

Tots els problemes apuntats fins aquí es veuen magnificats per la qualitat deficient de les publicacions, especialment per la manca de detalls que impedeixen la correcta valoració d'aspectes relacionats amb el risc de biaix o bé una correcta interpretació dels resultats. A l'àmbit de la lumbàlgia, especialment quan es tracta d'intervencions aplicades per col·lectius professionals no mèdics, les publicacions solen presentar més deficiències formals, en gran part perquè es tracta de revistes de menor difusió o no indexades. Moltes d'aquestes revistes no estan adherides a las recomanacions o estàndards internacionals sobre publicació d'assaigs clínics, com ara CONSORT⁽⁷⁰⁾ o la seva extensió per assaigs clínics d'intervencions no farmacològiques.⁽⁷¹⁾

Aquesta qüestió ja ha estat avaluada en una gran varietat d'àmbits clínics on s'empren diferents tipus d'intervencions no farmacològiques, i tots assenyalen que la qualitat dels estudis és subòptima i ha de ser millorada.⁽⁷²⁻⁸¹⁾ Segurament això és aplicable també a tot el rang d'intervencions no farmacològiques que es proposen per al tractament de la lumbàlgia. De fet, ha estat una tònica general a les nostres revisions les mancances en la descripció detallada d'aspectes importants dels mètodes d'estudi que sovint han impedit una avaluació més acurada del risc de biaix i que, en alguna mesura, han dificultat també la interpretació dels resultats. Això era més evident quan més antiga era la publicació. Com ja s'ha dit abans, aquest fet no és un problema de les revisions sinó dels estudis primaris però que, òbviament, limita la capacitat de les revisions de respondre les preguntes plantejades.

Limitacions i reptes metodològics per dur a terme assaigs clínics que avaluen intervencions no farmacològiques

Com ja s'ha apuntat abans, els assaigs clínics que formen part de les revisions sovint han presentat limitacions metodològiques. Segurament aquestes limitacions són comunes a gran part, si no a totes, les intervencions no farmacològiques que es fan servir per al tractament de la lumbàlgia i de qualsevol altra patologia.

Estudis que han comparat els AC d'intervencions no farmacològiques front a intervencions farmacològiques remarquen aquest mateix fet així com la necessitat de tenir en compte aspectes específics que afecten particularment als estudis que avaluen intervencions no farmacològiques.⁽⁷²⁾ Precisament això és el que va motivar la necessitat d'adaptar les directrius CONSORT a aquest tipus d'intervencions.⁽⁸²⁾

Algunes de les principals limitacions trobades en l'avaluació dels estudis inclosos a les quatre revisions presentades en aquesta tesi es comenten a continuació. La importància de tots aquests aspectes, i d'altres més, està molt ben explicada i raonada, amb exemples que ho il·lustren, en el document que

va servir de base per adaptar la proposta CONSORT als assaigs clínics d'intervencions no farmacològiques.⁽⁷¹⁾

Pel que fa a la població d'estudi:

- Sovint, la descripció de les característiques basals de la població d'estudi era incompleta i mancaven dades rellevants per entendre els resultats i la seva eventual aplicabilitat.
- De vegades, els criteris diagnòstics eren incomplets o imprecisos.
- Com en el cas de l'IDET, s'ha emprat una prova diagnòstica qüestionable per confirmar la presència de la patologia en estudi, la qual cosa pot afectar la selecció de la població d'estudi i també els resultats.
- D'altres vegades, tot i tractar-se d'una mateixa patologia, la població d'estudi barrejava diferents tipus de pacients (per exemple, pacients amb i sense dolor irradiat), amb un risc basal diferent, sense presentar dades de resultat per separat, la qual cosa impedeix fer una correcta interpretació dels resultats. La petita grandària dels estudis ha impedit fer anàlisis de subgrups que haurien resultat de molt interès.
- Com ha estat en el cas dels estudis SPORT sobre cirurgia per l'estenosi espinal, les expectatives dels pacients sobre els tractaments en investigació poden influir en la seva acceptació de l'estudi i, per tant, afectar la composició final de la mostra d'estudi i la seva representativitat. Donada la impossibilitat pràctica d'emascarar moltes intervencions no farmacològiques, això és un problema freqüent en estudis d'intervencions no farmacològiques, i més particularment en el cas de la cirurgia.
- A la majoria de casos, es tractava d'estudis de grandària petita. Això ha pogut provocar algun desequilibri a les característiques basals entre els grups, malgrat l'aleatorització. També ha impossibilitat fer anàlisis de subgrups.

Pel que fa a l'àmbit d'estudi:

- Alguns estudis no detallen el nombre i tipus de centre/s on s'aplica la intervenció i on es realitza el seguiment del pacient. Això és particularment important en aquelles intervencions el resultat de les quals depèn també de l'experiència (volum) i perfil de qui les aplica.
- Sovint, els estudis s'han portat a terme en un únic centre, per part d'investigadors amb conflictes d'interès.

Pel que fa a les intervencions:

- De vegades, les intervencions avaluades no estaven descrites amb prou detall, especialment les del grup control. Les intervencions no farmacològiques, per la seva pròpia naturalesa, solen ser força més complexes que un tractament amb medicaments i de vegades inclouen més d'un component. Cal, doncs, que els estudis descriguin amb prou detall les intervencions i la modalitat específica d'aplicació.
- L'elecció del grup control és un altre punt controvertit. El placebo (intervenció simulada o no fer res) seria el control ideal per demostrar eficàcia i un control actiu que hagi demostrat prèviament la seva eficàcia per demostrar superioritat. No obstant, sovint als estudis inclosos a les revisions s'ha seleccionat un control actiu quan encara no és prou clar si la intervenció és o no eficaç. Fins i tot, alguna vegada el control suposadament actiu consistia en una intervenció d'eficàcia no provada. Això condueix inevitablement a un conflicte en la interpretació dels resultats.
- Alguns estudis han apostat per un plantejament pragmàtic que deixa a la lliure elecció del metge responsable la modalitat de tractament que haurà d'aplicar en cada cas, segons les especificitats del pacient i altres condicionants locals (disponibilitat) o personals (experiència, preferències). Això ha estat present tant en algunes intervencions experimentals (per exemple, el tipus de cirurgia per l'estenosi espinal) com en el grup control (per exemple, tractament habitual). De vegades no és possible protocolitzar fins al darrer detall tots els aspectes de les

intervencions que es comparen, i aquesta variabilitat es fa inevitable, encara que planteja reptes de cara a la interpretació dels resultats. Novament, la petita grandària dels estudis ha dificultat fer anàlisis de subgrups o de regressió per identificar components de la intervenció que influeixen més en els resultats.

- Al contrari, quan els tractaments estaven estandarditzats, sovint no es descriuen els procediments seguits per tal de garantir l'aplicació d'un mateix protocol en tots els casos, ni tampoc es donen detalls del grau d'adherència dels terapeutes a aquest protocol.

Pel que fa a les variables de resultat:

- Algunes de les variables de resultat més habituals en aquest àmbit, com el dolor, la discapacitat i la funció, són mesures clínicament rellevants, amb instruments de mesura i criteris d'interpretació validats (VAS, Oswestry Disability Index scale, Roland Morris questionnaire, etc.). El consens internacional ha propiciat una major homogeneïtat entre estudis quant a les mesures de resultat emprades, la qual cosa facilita la seva interpretació i comparació entre estudis.^(83,61) No obstant, en els estudis de la revisió s'ha emprat una gran varietat de mesures de resultat addicionals a les abans esmentades, algunes d'elles creades *ad-hoc* (per tant, sense criteris d'interpretació validats).
- Moltes de les variables de resultat habituals en aquest àmbit consisteixen en variables subjectives reportades pels pacients. La manca d'emascament a la majoria d'estudis de les revisions planteja dubtes sobre un risc de biaix en la mesura dels resultats.
- De vegades, les mesures de resultat seleccionades no són clínicament rellevants, perquè no es correlacionen bé amb la situació de salut dels pacients.
- No sempre quedava clar quina era la mesura de resultat principal i quines les secundàries.

- La informació sobre els efectes adversos es percep com una àrea deficitària a la majoria dels estudis. No queda clar si la comunicació dels efectes adversos obeïa a una comunicació espontània per part del pacient o es tractava d'un registre sistematitzat i protocolitzat. Si bé és poc probable que els efectes adversos més greus passin desapercibuts, hi ha dubtes sobre la resta. És molt possible que hagin estat infraestimats, la qual cosa dificultarà la ponderació entre els beneficis i riscos que el metge haurà de fer juntament amb el seu pacient a l'hora de decidir el tractament que aplicarà en el context de la pràctica clínica.
- De vegades, la complexitat de la variable de resultat fa aconsellable garantir la qualitat de la seva medició, per exemple, amb més d'un avaluador o mitjançant avaluadors ensinistrats prèviament per fer la medició. Això no ha estat comunicat en gairebé cap estudi de les revisions.

Pel que fa al disseny de l'estudi:

- Aleatorització:
 - En molts casos, no es proporcionava detalls sobre el procediment d'aleatorització i, encara menys, sobre com garantir l'ocultació de la seqüència de les assignacions. Això ha plantejat dubtes sobre el risc d'un biaix de selecció.
 - En algun cas (revisió d'IDET), es tractava d'estudis controlats però no aleatoris, que varen proporcionar resultats favorables i en un sentit oposat als estudis de major qualitat. Probablement es tractava d'una mostra esbiaixada. Els estudis SPORT també inclouen una cohort prospectiva de naturalesa observacional que no es va tenir en compte a la revisió per aquesta mateixa raó.
- Emmascarament:
 - Alguns estudis han fet esforços importants per tal d'emascarar les intervencions (per exemple, la NRT simulada). D'altres, no ho

han fet i, encara en uns altres, això era del tot impossible (cirurgia de l'estenosi espinal).

- En el primer cas, és molt important descriure el procediment per tal de garantir l'èxit de l'emascament. Per exemple, als estudis de NRT versus simulació hi ha una àmplia informació al respecte, mentre que en altres casos gairebé no es menciona res.
- Quant l'emascament no és possible, com és el cas de moltes intervencions no farmacològiques i en funció del comparador, cal intentar que l'avaluació de resultats sigui feta de forma cega. Als estudis SPORT això no es va fer per raons pràctiques de viabilitat, la qual cosa resta algun grau de credibilitat als seus resultats.
- Seguiment dels pacients:
 - La majoria d'estudis han portat a terme un seguiment a curt termini, malgrat tractar-se d'una patologia crònica i/o recurrent. Això dona una perspectiva curta sobre el valor de les intervencions, ja que no permet saber quant temps es manté l'efecte o si els pacients recauen. També impedeix conèixer bé el perfil de seguretat. Alguns estudis, una vegada assolit l'objectiu principal, han portat a terme una fase d'extensió a llarg termini de naturalesa observacional (només els pacients del grup d'intervenció, com va ser el cas de la cirurgia o la NRT).
 - En alguns casos, hi ha hagut un important nombre de creuaments (pacients que canvien de grup), com ja s'ha comentat en la revisió de la cirurgia per l'estenosi espinal. Això planteja reptes per l'anàlisi i interpretació dels resultats (anàlisi segons tractament rebut).
 - No sempre quedava clar el nombre d'abandonaments i pèrdues en el seguiment i com han estat comptabilitzades a l'anàlisi. En part, aquest problema ve provocat per l'absència d'un flux de pacients al llarg de l'estudi, tal como recomana CONSORT, a

molts dels estudis. Hi ha dubtes que tots els estudis hagin aplicat correctament una anàlisi per intenció de tractar, tot i que diguessin haver-lo aplicat (gairebé mai es donen detalls sobre les assumpcions dels autors). En alguns casos, directament s'ha optat per l'anàlisi per protocol que podria haver esbiaixat els resultats.

- Aspectes estadístics i d'anàlisi:
 - Molt estudis no justificaven el càlcul de la grandària mostral, impedit conèixer les assumpcions de partida, útils per interpretar els resultats.
 - Gairebé mai s'especificava la diferència mínima clínicament rellevant.
 - La majoria d'estudis eren de grandària petita, la qual cosa impossibilitava fer anàlisis de subgrups, cosa que hauria resultat interessant en alguns casos (per exemple, per identificar subgrups de pacients que més es beneficien del tractament o modalitats específiques de tractament que presenten millors resultats). Tampoc ha permès fer ajustos per controlar variables confonedores.
 - En alguns estudis, el criteri per establir l'èxit del procediment era del tot arbitrari o fins i tot es va establir *a posteriori*.
 - Perques vegades es comentava la rellevància clínica dels resultats. En alguns casos, l'únic criteri que ha guiat la interpretació dels resultats ha estat el valor de la p.
 - Com ja s'ha dit, sovint hi havia força confusió sobre el tipus d'anàlisi de resultats aplicada. De vegades, no quedava clar el nombre de pacients analitzat ni tampoc si s'havien fet o no imputacions dels valors no disponibles (*missings*). La manera de presentar els resultats a les taules sovint no ha ajudat a resoldre aquest problema.

- Aspectes ètics:
 - Es percep que en molts casos hi ha potencials conflictes d'interès, ja sigui de tipus econòmic o professional. A molts estudis no es mencionava res al respecte.

En conclusió, cal aprofundir en el coneixement metodològic dels aspectes clau que afecten els estudis sobre intervencions no farmacològiques, i sobre tot, cal seguir fomentant (i exigint) que es facin aquests estudis amb el màxim rigor científic. Per tal que això sigui possible, caldrà també que des del sistema de salut es donin facilitats (finançament, formació, estructures de suport a la recerca, etc.) per fer que la recerca en aquest àmbit sigui possible i de qualitat. Només així, el nivell d'evidència disponible serà prou alt per poder orientar millor les decisions clíniques i sanitàries en una patologia on encara hi ha moltes incerteses.

[Deficiències en el procés d'aprovació i introducció d'intervencions no farmacològiques per a les patologies de l'esquena al sistema de salut](#)

Aquest darrer apartat de la tesi doctoral pretén ser la culminació de tota la reflexió aportada als apartats anteriors. La tesi subjacent és ben simple i evident: si no es disposa d'informació de qualitat, difícilment es poden prendre decisions clíniques i sanitàries que fomentin un ús racional de les intervencions terapèutiques disponibles. Ja s'ha apuntat abans la necessitat de promoure la recerca en aquest àmbit perquè es generi la informació que cobreixi les llacunes del coneixement existents sobre el tractament de les patologies de l'esquena, però també cal posar en marxa mecanismes més rigorosos i eficients que vetllin perquè, amb la informació disponible, es prenguin les decisions més adequades, en benefici de la salut dels pacients i també de la sostenibilitat i eficiència dels sistemes de salut.

Els mecanismes que condueixen a l'aprovació (autorització) i introducció de les intervencions terapèutiques al sistema de salut són manifestament millorables,

fins i tot quan es tracta de dispositius i productes sanitaris.⁽⁸⁴⁻⁸⁸⁾ La feble evidència (de vegades absència) en què es basen moltes de les intervencions i procediments habituals en el sistema sanitari, així com la manca d'uns resultats clarament favorables en termes de salut, així ho indiquen.^(47,89) La creixent i accelerada difusió de tecnologies mèdiques de tota mena per a la prevenció i tractament de patologies de l'esquena a la pràctica clínica suposa un important repte per a la capacitat del sistema d'analitzar críticament i de forma imparcial l'eficàcia, seguretat i efectivitat de totes i cadascuna d'aquestes intervencions i per fer-ne un seguiment una vegada ja han estat introduïdes al sistema sanitari.⁽⁸⁹⁾

La necessitat de millorar aquests mecanismes implica no només a les agències d'avaluació de tecnologia mèdica (que amb molta probabilitat ja fan bé la seva feina pel que fa a les tecnologies més complexes), sinó també i sobretot als organismes reguladors, als col·lectius professionals (especialment les societats científiques implicades en l'elaboració de guies de pràctica clínica i documents de consens) i als responsables o gestors dels recursos sanitaris als diferents àmbits del sistema de salut.

Lluny d'assenyalar les responsabilitats de cadascun (aquest no és el propòsit de la tesi), sí cal fer una reflexió general sobre les deficiències del marc que governa el procés d'avaluació i introducció de les intervencions terapèutiques no farmacològiques en l'àmbit de les patologies de l'esquena. Si només pensem en les intervencions que impliquen l'ús de tecnologia sofisticada o dispositius costosos als que estem acostumats en els hospitals terciaris, segurament es tracta d'un entorn força regulat, encara que amb insuficiències, com demostra la recent polèmica en torn a l'ús de les proteïnes morfogenètiques de l'òs i la seva seguretat.⁽⁵³⁻⁵⁴⁾ Però si per tecnologia entenem tota intervenció, independentment de la seva complexitat i costos, fins i tot quan la tecnologia implicada consisteix essencialment en les mans del terapeuta que les aplica (cirurgià, psicòleg, fisioterapeuta, etc.), llavors trobarem àmplies àrees de desgovern. La variabilitat observada a la pràctica clínica en seria un clar indicador.⁽⁹⁰⁻⁹²⁾

Algunes de les intervencions que han estat objecte de les revisions presentades a la tesi ajuden a il·lustrar bé aquesta problema.

L'exemple de l'IDET

L'IDET va ser inicialment desenvolupada als EEUU a mitjans dels anys 90 pels germans Saal, co-fundadors d'Oratec, una companyia dedicada a desenvolupar i comercialitzar dispositius mèdics que utilitzen l'energia tèrmica per al tractament de patologies del raquis i articulars. Al març de 1997, després de 18 mesos de portar a terme diversos estudis clínics, cap dels quals es va publicar a una revista científica indexada i rigorosa, Oratec va presentar el dispositiu SpineCath per dur a terme el procediment IDET (electrotermocoagulació intradiscal percutània) al congrés de l'Acadèmia Americana de Cirurgians Ortopèdics.

Al març de 1998 Oratec va rebre l'autorització de la FDA (Food and Drug Administration, l'agència reguladora americana) per a la comercialització de SpineCath per al tractament de la lumbàlgia simptomàtica en pacients amb disrupció de l'anell fibrós del disc intervertebral, amb hèrnia discal continguda.⁽⁹³⁾ L'aprovació es va obtenir a partir de les dades preliminars (no publicades) d'un únic estudi en només 25 pacients. La FDA va atorgar l'aprovació de SpineCath en absència d'un estudi aleatori controlat que demostrés l'eficàcia d'IDET versus placebo, gràcies a que ja estava comercialitzat prèviament un altre dispositiu de característiques similars. La regulació vigent als EEUU (Medical Devices Act) estableix un procediment d'aprovació (conegut com a 510(k)) que només requereix la demostració d'una "equivalència substancial" (en el sentit de sistema de funcionament del dispositiu i no de resultats clínics) entre el nou dispositiu i un altre ja comercialitzat abans del 1976.^(89,94) En conseqüència, la FDA mai va avaluar l'eficàcia i seguretat de SpineCath, tot i que els mateixos autors reconeixien que *"les conseqüències a llarg termini de l'IDET amb aquest catèter son actualment desconegudes"*.

La taula 4 (procedent de l'article que s'adjunta en l'annex 5) presenta alguns dels mites sobre el procés d'aprovació i control dels nous dispositius i que, sovint, porten a la creença generalitzada de que quan un dispositiu obté autorització per a ser comercialitzat això suposa un aval de la seva eficàcia i seguretat clínica.⁽⁸⁹⁾ Encara que la taula fa referència a la FDA (Food and Drug Administration, l'agència reguladora dels EEUU), les afirmacions que s'hi fan són igualment vàlides per Europa on el procés d'aprovació és encara menys rigorós.^(84,86,88)

Taula 4. Mites i realitats sobre l'avaluació i regulació dels dispositius mèdics

Mite	Realitat
El nou dispositiu ha de ser millor que un altre de vell per ser aprovat.	Per a l'aprovació d'un nou dispositiu, no es requereix la seva comparació amb els dispositius de la competència. Un nou dispositiu pot ser igualment eficaç, més eficaç o menys eficaç que les alternatives.
Els estudis realitzats per a la seva aprovació és probable que hagin detectat tots els efectes secundaris importants.	Pels nous fàrmacs, el 50% dels efectes secundaris importants són descoberts després de la seva aprovació i comercialització. No hi ha dades disponibles per als dispositius, però les llacunes en el procés d'aprovació suggereixen que amb molta probabilitat el problema és almenys d'una magnitud similar.
El procés d'aprovació dels dispositius és similar al dels medicaments.	La majoria dels dispositius (més del 98%) són aprovats amb la demostració d'una "equivalència substancial" amb algun altre dispositiu que ja està en el mercat. Això pot semblar paradoxal, ja que el <i>marketing</i> en general fa èmfasi en la novetat i la innovació. L'avaluació es centra en el funcionament i en la seguretat, més que no pas en l'eficàcia clínica. A diferència dels medicaments, no es requereixen assaigs clínics, excepte per als dispositius considerats com els més innovadors i arriscats.

Els dispositius són aprovats només si ofereixen una relació qualitat-preu favorable.	La FDA té prohibit considerar el cost o el cost-efectivitat en el procés d'aprovació.
Els usos fora de indicació dels dispositius són ben avaluats.	L'ús fora de indicació de molts dispositius és una pràctica habitual i legal, però la seva eficàcia i seguretat sovint no estan ben establertes perquè no hi ha cap incentiu per estudiar-ho una vegada el dispositiu ja ha estat aprovat com a mínim en una indicació.
Els dispositius ineficaços són revisats.	Els dispositius són revisats només quan hi ha problemes de seguretat i no per l'acumulació de proves de la seva manca d'eficàcia.
Els procediments quirúrgics estan regulats per la FDA.	La FDA no regula cap procediment, només dispositius.

A l'octubre d'aquell mateix any es produeix el llançament del producte a la conferència de la North American Spine Society (NASS). Des de llavors, s'estima que s'han dut a terme 75.000 procediments d'IDET als EEUU.⁽⁹⁵⁾ La seva ràpida disseminació va començar fins i tot abans de la publicació del primer estudi clínic en una revista indexada d'impacte científic l'any 2000.⁽⁹⁶⁾ Aquest estudi (presentat com a estudi preliminar al seu títol) representa la primera comunicació de resultats amb IDET en només 25 pacients. Posteriorment, es va publicar una gran quantitat de sèries de casos (estudis observacionals, sense grup control, en disseny obert) per part dels germans Saal i col·laboradors, en revistes de baixa difusió i impacte, totes elles comunicant resultats altament favorables a IDET i sense cap risc associat.

Al febrer de 2002, Oratec és comprada per una gran companyia multinacional (Smith&Nephew), una decisió motivada per les bones perspectives comercials, segons la informació obtinguda al web de la companyia: *"It is expected that IDET sales (\$21m in 2001) will recover and grow following the publication of results of two-year follow up clinical studies in peer-reviewed publications"*.⁽⁹⁷⁾

Al 2004 es publica el primer assaig clínic, promogut per la companyia, a una revista indexada⁽⁹⁸⁾, que comparava IDET versus placebo (simulació), amb

resultats favorables a IDET per al dolor i la discapacitat. No obstant, aquest estudi presentava importants problemes metodològics que qüestionen els seus resultats, a més de la seva dubtosa rellevància clínica. L'any 2006, un segon AC independent ben dissenyat comparant IDET versus placebo conclou amb resultats desfavorables a l'IDET.⁽⁹⁹⁾ Tal i com assenyala la nostra revisió, *“en conjunt, l'evidència disponible és contradictòria, i no permet arribar a una conclusió ferma sobre l'eficàcia d'IDET”*. Aquesta discrepància freqüent entre els resultats obtinguts per estudis promoguts per la indústria i els independents porta a plantejar-se el paper i la veritable independència dels investigadors acadèmics (sovint líders d'opinió) que presten el seu nom per dur a terme investigacions amb clars interessos comercials. Els conflictes d'interès, molt sovint no declarats en aquest àmbit de les patologies de l'esquena, qüestionen la integritat de la recerca.⁽⁸⁹⁾

Quant al paper que, sovint, juguen els col·lectius professionals i les societats científiques, el cas de l'IDET és revelador. Així, l'American Society of Interventional Pain Physicians (ASIPP), creada l'any 1998 amb l'objectiu de *“promoure el desenvolupament i l'ús d'intervencions segures, de qualitat i cost-efectives”* va publicar l'any 2007 unes guies que conclouen que *“l'evidència per al tractament de la lumbàlgia crònica discogènica amb IDET es moderada”*.⁽¹⁰⁰⁾ No obstant això, als seus algorismes, la guia recomana explícitament l'ús d'IDET. Davant aquesta contradicció, sorgeix la sospita de que hi podria haver conflictes d'interès que podrien influir en els judicis i decisions dels autors de la guia, molt d'ells implicats en l'ús d'aquesta tècnica. L'any 2008, al web d'aquesta societat apareixia la companyia Smith&Nephew com un dels principals patrocinadors (categoria d'or).

Un cas oposat seria el de la guia europea COST B13, realitzada de forma completament independent de la indústria, i que conclou que *“l'evidència sobre la superioritat d'IDET versus placebo és contradictòria”*, i per tant, *“no es recomana el seu ús”*.⁽¹⁾ Mateixa evidència, diferents judicis i conclusions oposades. Així mateix, la majoria d'agències d'avaluació de tecnologia mèdica no recolzen l'ús d'IDET i jutgen les evidències com a “no convincents”, “insuficients”, “inadequades” o “dèbils” (taula 5).

Taula 5. Conclusions d'alguns informes d'avaluació de tecnologia sobre l'IDET.

Institute for Clinical Systems Improvement (2002) ⁽¹⁰¹⁾

“There is no convincing evidence that shows the short or long-term clinical efficacy of this procedure”.

Medical Service Advisory Committee (Australia) (2003) ⁽¹⁰²⁾

“Since there is currently insufficient evidence pertaining to intradiscal electrothermal annuloplasty... MSAC recommended that public funding should not be supported at this time for this procedure”.

Danish Center for Evaluation and Health Technology Assessment (2003) ⁽¹⁰³⁾

“The literature does not yet provide a clear picture of IDET’s clinical effect... As there is no convincing documentation for the indications and treatment results, in the event that IDET is introduced in Denmark this should take place via a randomised controlled clinical trial”.

National Institute for Clinical Excellence (NICE) (2004) ⁽¹⁰⁴⁾

“Current evidence on the safety and efficacy of percutaneous intradiscal electrothermal therapy for lower back pain does not appear adequate to support the use of this procedure without special arrangements for consent and for audit or research”.

Montreal: Agence d’Evaluation des Technologies et des Modes d’Intervention en Sante (AETMIS) (2005) ⁽¹⁰⁵⁾

“Given the weakness of the evidence, a number of assessments recommend not including intradiscal electrothermal therapy as an insured service, except in a research setting”.

The BlueCross / BlueShield Association Technology Evaluation Center (2007) ⁽¹⁰⁶⁾

“Intradiscal electrothermal annuloplasty or percutaneous intradiscal radiofrequency thermocoagulation is considered investigational as a treatment of chronic discogenic back pain”.

Aquest exemple vol il·lustrar les contradiccions en el paper que hi juguen els diversos actors implicats on, sovint, l'entusiasme dels professionals (i també del públic) per les novetats tecnològiques no es correspon amb les evidències disponibles sobre la seva utilitat. Segurament això és comú a moltes altres tecnologies emprades per al tractament de les patologies de l'esquena, com mostren els estudis ja mencionats sobre variabilitat en les taxes de cirurgia

d'esquena als EEUU (especialment l'ús d'atrodesi instrumentada)⁽⁴⁷⁾ i també quant a l'ús dels ultrasons.

Sovint s'observa un mateix patró:

- Una nova tecnologia es desenvolupa, habitualment amb clars interessos comercials (quan es tracta de dispositius) i/o professionals. Òbviament, els interessos comercials o professionals no constitueixen *per se* un problema que invalidi els resultats ni tampoc la rellevància de la innovació, però és un factor a tenir molt en compte i que hauria d'obligar a una total transparència en la declaració dels conflictes d'interessos.⁽¹⁰⁷⁾
- Experts (sovint líders d'opinió) promouen l'ús de la tecnologia en reunions científiques presentant resultats anunciats com a interessants o impactants sorgits d'estudis de qualitat limitada (com sèries de casos) que, com a molt, haurien de ser considerats com a evidència preliminar però mai com a una evidència definitiva o confirmatòria.⁽¹⁰⁸⁾
- S'obté l'autorització per la comercialització del procediment o dispositiu per part de les agències reguladores sense massa dificultat, ja que el llindar d'exigències sol ser molt més baix i el procés és menys rigorós que en el cas dels fàrmacs.⁽⁸⁹⁾
- Sovint, la companyia implicada posa en marxa una política de publicacions pseudocientífiques consistent en múltiples articles, de baixa qualitat i/o redundants (per exemple, duplicació, publicacions parcials, anàlisis de subgrups *post-hoc*, etc.)⁽¹⁰⁹⁻¹¹¹⁾, sovint escrits per autors fantasma⁽¹¹²⁾, amb resultats molt prometedors, que són publicats a revistes de baixa qualitat (suplements, procés de revisió poc exigent o fins i tot inexistent en revistes de pagament), que contribueixen a crear un estat d'opinió favorable que promou la disseminació de la tecnologia.
- Els mitjans de comunicació generals i especialitzats contribueixen a generar una expectativa, que col·loca una pressió creixent en els decisors de salut.⁽¹¹³⁻¹¹⁷⁾

- Com a conseqüència, es produeix una àmplia difusió de la tecnologia o la intervenció no farmacològica malgrat l'absència de (suficient) evidència sobre la seva eficàcia i seguretat, no només al sistema privat sinó de vegades també al públic.
- Algunes societats científiques sovint juguen un paper que contribueix a generar més confusió, en la mesura que tenen conflictes d'interès o bé estan orientades a la defensa dels interessos corporatius per damunt de la promoció d'una medicina basada en l'evidència.⁽¹¹⁸⁻¹²⁰⁾
- En aquesta atmosfera favorable, el promotor promou un ús abusiu de la tecnologia ampliant el rang de les seves indicacions (*off label*).⁽¹²¹⁾
- Arribats a aquest punt d'intoxicació informativa, pocs s'atreveixen a qüestionar l'ús de la tecnologia o de la intervenció novedosa, malgrat la manca de dades fiables sobre la seva eficàcia i seguretat (també a llarg termini) i cost-efectivitat, de vegades fins que es produeix un problema greu (com ara ha estat el cas de les proteïnes morfogenètiques de l'òs o abans ho varen ser les pròtesis de disc intervertebral)^(53,54).
- El paper i la influència de les agències d'avaluació de tecnologia, de vegades, es veu limitat en la mesura que els responsables polítics estan sotmesos a pressions externes (per part de la indústria, els lobbies mèdics o els mitjans de comunicació) que acaben influint en les decisions sobre finançament i prestació de serveis.
- Com a resultat, molts tractaments inútils, en el pitjor dels casos fins i tot nocius, o en el millor dels casos que estan encara en una fase investigacional, són àmpliament disseminats i pagats amb recursos públics.
- Finalment, tot això pot conduir a fomentar una pràctica clínica inadequada, a una despesa ineficient i/o a una exposició dels pacients a riscos innecessaris.

L'exemple de la NRT

Pel contrari, la neuroreflexoteràpia (NRT) il·lustra quin hauria de ser el procés d'avaluació i implantació d'una tecnologia sanitària al sistema de salut.

De forma resumida, es poden identificar diverses etapes que se segueixen en un ordre seqüencial i lògic.

1.- Generació d'evidència científica:

- a. desenvolupament inicial de la tècnica als anys 80 a partir d'estudis de neurofisiologia del dolor (establiment de les bases biològiques del procediment i de seu fonament teòric).
- b. inici d'estudis clínics pels seus impulsors (Fundación Kovacs).
- c. primera publicació d'una àmplia cohort prospectiva (2.571 pacients) que proporciona evidència preliminar de la seva eficàcia i seguretat.⁽¹²²⁾
- d. publicació d'un primer assaig clínic (AC) de qualitat que proporciona una sòlida evidència sobre la seva eficàcia (comparació de NRT versus placebo).⁽¹²³⁾
- e. replicació dels resultats mitjançant un segon AC de característiques similars al primer.⁽¹²⁴⁾
- f. publicació d'un tercer AC de filosofia pragmàtica que proporciona evidència sobre la seva superioritat respecte el maneig habitual a l'atenció primària i el cost-efectivitat també a llarg termini (1 any) en un context de la pràctica real.⁽¹²⁵⁾
- g. síntesi de l'evidència disponible mitjançant revisions sistemàtiques⁽¹²⁶⁾, informes de tecnologia⁽¹²⁷⁻¹³⁰⁾ o tècnics⁽¹³¹⁾ i guies de pràctica clínica⁽¹⁾ (investigació secundària).

2.- Implementació a la pràctica clínica:

- h. suport racional a la decisió d'introduir la NRT al sistema de salut de les illes balears (Ib-Salut) a partir de la investigació secundària (informes de síntesi).
- i. planificació del procés d'implementació de la NRT a la practica clínica (atenció primària), amb la definició d'estàndards (criteris d'indicació, protocol de derivació, protocol d'aplicació i avaluació prospectiva de resultats de la NRT).⁽¹³²⁻¹³⁴⁾
- j. experiència pilot - estudi d'avaluació dels estàndards (factibilitat, satisfacció i resultats clínics preliminars) amb participació de 18 centres de primària i 208 metges (2002).⁽¹³⁵⁾
- k. implantació de la NRT a la pràctica clínica i control per l'Ib-Salut.
- l. vigilància *post-marketing* i confirmació dels resultats preliminars a llarg termini obtinguts als AC a la pràctica clínica real.⁽⁶²⁾
- m. optimització dels criteris d'indicació i dels resultats clínics (mitjançant l'estudi dels factors pronòstics i predictius de bons resultats de la NRT).⁽¹³⁶⁾
- n. difusió de la NRT a d'altres àrees o sistemes de salut de l'estat espanyol (Astúries, Múrcia, Madrid, prova pilot a Catalunya).

Com a conclusió, cal implantar un model més rigorós d'avaluació i implantació de les tecnologies i intervencions no farmacològiques pel tractament de les patologies de l'esquena en el sistema de salut que impliqui a totes les parts interessades on:

1. Es promogui la recerca clínica (idealment conduïda o amb participació d'organismes independents i, si és possible, amb finançament públic segons la seva rellevància pel sistema de salut) pertinent i de qualitat (assaigs clínics, amb emmascarament de les intervencions o cegament de l'avaluació de resultats, amb un grup control adequat, variables de

resultat rellevants, seguiment a llarg termini, mida àmplia...) que permeti generar evidència fiable i de bona qualitat per contestar les preguntes:

- i. És segura?
- ii. És millor que placebo?
- iii. És millor que altres tractaments disponibles amb eficàcia provada?
- iv. És cost-efectiva?

Cal que es garanteixi la transparència en la difusió dels resultats (evitant el biaix de publicació i de comunicació selectiva de resultats) mitjançant el registre prospectiu dels assaigs clínics, així com una publicació de qualitat (seguint recomanacions internacionals com CONSORT) per facilitar la correcta utilització d'aquesta informació pels seus usuaris finals (els decisors).

Així mateix, cal que els investigadors adoptin una actitud més rigorosa, implicant-se més activament en el disseny dels estudis per assegurar que aquests siguin clínicament rellevants, participant en les decisions sobre la publicació de resultats, evitant l'autoria per part d'autors fantasma, evitant estratègies de publicació fraudulentos o esbiaixades, assegurant l'accés sense restriccions a les dades i supervisant la seva anàlisi segons l'establert prèviament al protocol.

2. En la mesura que es generin i publiquin estudis científics sobre la tecnologia o la nova intervenció, cal portar a terme revisions sistemàtiques que avaluïn de manera crítica i independent tot aquest cos d'evidència des d'una perspectiva global i integradora per tal de concloure sobre la seva eficàcia i seguretat. Aquestes revisions haurien de seguir també els estàndars internacionals de publicació (PRISMA).⁽¹³⁷⁾

Pel que fa a les agències reguladores que han d'emprar la informació proporcionada per les revisions sistemàtiques i estudis originals per prendre les seves decisions, haurien d'actuar amb completa

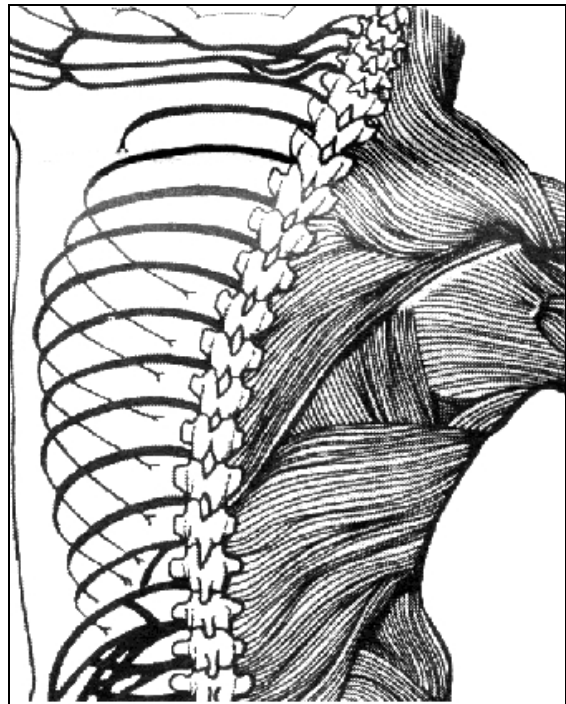
independència, lliures de pressions externes, per tal de garantir que els conflictes d'interès no interfereixen amb una adequada valoració no esbiaixada de la utilitat de les tecnologies.

3. L'administració sanitària i/o les societats científiques (idealment ambdues conjuntament) haurien d'elaborar guies de pràctica clínica basades en l'evidència, contextualitzant la informació procedent de les revisions i dels estudis primaris (en absència d'aquelles), per tal de concloure amb recomanacions explícites, favorables o no, a l'ús de la tecnologia o intervenció en qüestió. Les societats científiques (o els experts que hi participen en les guies) haurien de ser transparents en la declaració dels seus conflictes d'interès.
4. Si s'escau, s'hauria de planificar la seva introducció a la pràctica clínica, definint prèviament les indicacions concretes i les modalitats o formes específiques d'aplicació que garanteixin l'efectivitat i seguretat del procediment o intervenció.
5. Un cop s'ha decidit la seva introducció, caldria portar a terme un estudi pilot per comprovar que la tecnologia s'aplica correctament i es monitoritzen els resultats, que hauran de servir per prendre la decisió de continuar o no endavant amb la prestació (si cal, amb una autorització condicionada).
6. Una vegada introduïda i difosa, el sistema de salut hauria d'auditar periòdicament els resultats (vigilància postmarketing) per tal de refinar els criteris d'indicació, optimitzar els resultats clínics (taxa d'èxit) i garantir la seguretat de la tecnologia. Els hospitals i els professionals també hi tenen alguna responsabilitat, en la mesura que haurien de ser capaços de mesurar els seus resultats clínics, especialment quan es tracta d'intervencions complexes, costoses i potencialment arriscades.
7. Només llavors hi hauria la base suficient per a una progressiva generalització del seu ús dins el sistema.

Per tal que això sigui possible, cal la implicació de tots els actors (indústria, proveïdors/finançadors, investigadors i clínics) amb uns mateixos objectius de

promoure la salut dels pacients, un compromís clar pel rigor i la transparència i, sobretot, un sistema nacional de salut que actuï amb rigor, fermesa i de forma proactiva per garantir l'assemblatge de tot l'engranatge i que els interessos dels pacients i socials prevaleixin sobre els particulars i comercials.

7. Conclusions



7.- CONCLUSIONS

7.1.- Implicacions per a la pràctica

1. Són nombrosos els tractaments i intervencions no farmacològiques que actualment es proposen als pacients amb patologies mecàniques de l'esquena, moltes de les quals no estan suficientment avalades per estudis científics de qualitat que hagin demostrat la seva eficàcia, efectivitat i seguretat.
2. De les intervencions avaluades a la tesi, dues han demostrat ser efectives i segures i es poden recomanar: la NRT per al tractament de la lumbàlgia crònica i subaguda no específica, i la cirurgia de descompressió (amb o sense fusió) en determinats pacients de risc amb estenosi espinal simptomàtica (sobretot en presència de claudicació neurògena) que han fracassat a un tractament conservador previ després de 3 a 6 mesos.
3. L'evidència sobre l'efectivitat i seguretat de les tècniques percutànies d'electrotermocoagulació intradiscal (IDET) per al tractament de la lumbàlgia discogènica és encara no conclouent i, per tant, no es pot recomanar.
4. Els ultrasons o les ones de xoc no semblen tenir cap efecte favorable clínicament rellevant per al tractament de la lumbàlgia.

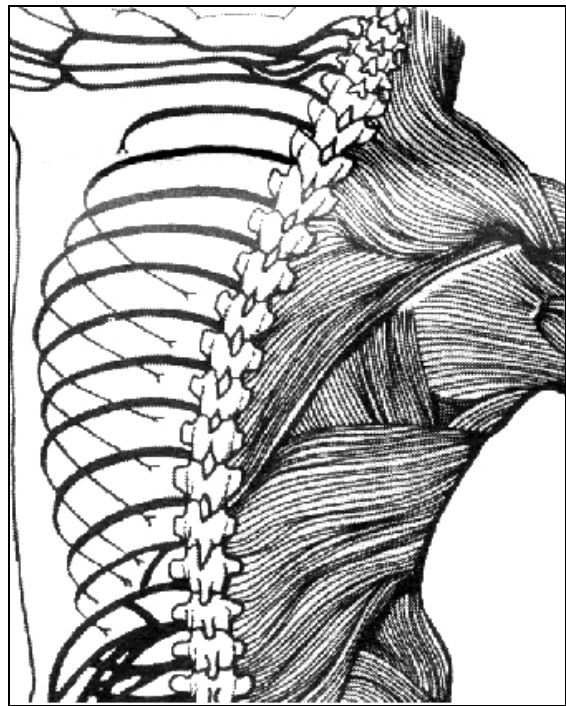
7.2.- Implicacions per a la recerca

1. Cal seguir investigant en alguns aspectes relacionats amb les intervencions avaluades per les revisions com, per exemple: 1) el tipus específic de cirurgia més efectiva (especialment el paper de la fusió instrumentada) i els factors pronòstics dels pacients amb estenosi espinal (refinament dels criteris d'indicació), 2) seguretat i confirmació dels resultats a llarg termini de la cirurgia, 3) confirmació dels resultats del dispositiu interespinal de tracció amb nous estudis (replicació), 4) cal

més investigació sobre procediments i tècniques mínimament invasives, com l'IDET.

2. Les agències de finançament públic de la recerca haurien d'incrementar la seva inversió en l'àrea de les patologies de l'esquena i també revisar i millorar els seus mecanismes d'avaluació i selecció dels projectes, que haurien de ser més transparents.
3. Cal promoure la realització d'assaigs clínics d'intervencions no farmacològiques en l'àrea de les patologies de l'esquena (cirurgia, fisioteràpia, teràpia física, intervencions invasives, educacionals o psicològiques, etc.) creant una cultura de l'avaluació i facilitant els mitjans per fer-ho. Així mateix, cal millorar la qualitat dels estudis i assaigs clínics en aquesta àrea, superant les nombroses mancances detectades, així com també la de les publicacions derivades, idealment en revistes indexades, seguint les recomanacions de la declaració CONSORT.

8. Bibliografia



8.- BIBLIOGRAFIA

[1] European Commission COST B13, Fundació Kovacs. Guia de pràctica clínica para la lumbalgia inespecífica; 2005. Disponible a: <http://www.webdelaespalda.org/cientifica/guias/guiascost.asp> [consultat el 10 de juny de 2011].

[2] Chou R, Qaseem A, Snow V, Casey D, Cross JT, Shekelle P, et al. Diagnosis and treatment of low back pain: A joint clinical practice guideline from the American College of Physicians and the American Pain Society. *Ann Intern Med* 2007;147:478-91.

[3] Deyo RA, Weinstein JN. Low back pain. *N Engl J Med* 2001;344:363-70.

[4] Waddell G. *The Back Pain Revolution*. Edinburgh: Churchill Livingstone, 1998.

[5] Chen SM, Liu MF, Cook J, et al. Sedentary lifestyle as a risk factor for low back pain: a systematic review. *Int Arch Occup Environ Health* 2009;82(7):797-806.

[6] Frymoyer JW, Pope MH, Constanza MC, Rasen J.C, Gogging G, Wildder DG. Epidemiologic studys of low back pain. *Spine* 1980;5:419-23.

[7] Deyo RA, Tsui-Wu JV. Functional disability due to low-bak pain. *Arthritis and Rheumatism* 1987;30:1243-7.

[8] Bergenudd H, Nilson B. Back pain in middle age; Occupational work load and psychologic factors. An epidemiologic survey. *Spine* 1988;131:58-60.

[9] Lundberg U, Granquist M, Hansson T, Magnusson M, Wallin L. Psychological and physiological stress responses during repetitive work at an assembly line. *Work and Stress* 1989;2:143-53.

[10] Thorbjornsson CO, Alfredsson L, Frediksson K, Koster M, Michelsen H, Vingard E, Torgen M, Kilbom A. Psychosocial and physical risk factors associated with low back pain: a 24 year follow up among women and men in a broad range occupations. *Occup Environ Med* 1998;55:84-90.

- [11] Gluck JV, Oleinick A. Claim rates of compensable back injuries by age, gender, occupation, and industry. Do they relate to return-to-work experience? *Spine* 1998;23:1572-8.
- [12] Haldorsen EMH, Indahl A, Ursin H. Patients with low back pain not returning to work. A 12-month follow-up study. *Spine* 1998;23:1202-8.
- [13] Devereux JJ, Buckle PW, Vlachonikolis IG. Interactions between physical and psychosocial risk factors at work increase the risk of back disorders: an epidemiological approach. *Occupa Environ Med* 1999;56:343-53.
- [14] Manchikanti L, Singh V, Datta S, Cohen SP, Hirsch JA. Comprehensive Review of Epidemiology, Scope, and Impact of Spinal Pain. *Pain Physician* 2009;12:E35-E70.
- [15] Kent PM, Keating J. The epidemiology of low back pain in primary care. *Chiropractic & Osteopathy* 2005,13:13 (doi:10.1186/1746-1340-13-13).
- [16] Walker BF, Muller R, Grant WD: Low back pain in Australian adults. Prevalence and associated disability. *Journal of Manipulative & Physiological Therapeutics* 2004;27(4):238-44.
- [17] NICE clinical guideline 88. Savigny P, Kuntze S, Watson P, Underwood M, Ritchie G , Cotterell M, Hill D, Browne N, Buchanan E, Coffey P, Dixon P, Drummond C, Flanagan M, Greenough,C, Griffiths M, Halliday-Bell J, Hettinga D, Vogel S, Walsh D. Low Back Pain: early management of persistent non-specific low back pain. London: National Collaborating Centre for Primary Care and Royal College of General Practitioners (2009).
- [18] Walker BF. The prevalence of low back pain: a systematic review of the literature from 1966 to 1998. *J Spinal Disord* 2000;13(3):205-17.
- [19] Lawrence RC, Helmick CG, Arnett FC. Estimates of the prevalence of arthritis and selected musculoskeletal disorders in the United States. *Arthritis Rheum* 1998;41:778-99.

- [20] Harkness EF, Macfarlane GJ, Silman AJ, McBeth J. Is musculoskeletal pain more common now than 40 years ago?: Two population-based cross-sectional studies. *Rheumatology (Oxford)* 2005;44:890-5.
- [21] Humbría A, Carmona L, Peña JL, Ortiz AM. Impacto poblacional del dolor lumbar en España: resultados del estudio EPISER. *Rev Esp Reumatol* 2002;29(10):471-8.
- [22] Nachemson AL. Newest knowledge of low back pain. A critical look. *Clin Orthop* 1992;279:8-20.
- [23] Pengel LH, Herbert RD, Maher CG, Refshauge KM. Acute low back pain: systematic review of its prognosis. *BMJ* 2003;327(7410):323.
- [24] Hestbaek L, Leboeuf-Y de C, Manniche C. Low back pain: what is the long-term course? A review of studies of general patient populations. *Eur Spine J* 2003;12(2):149-65.
- [25] Van Tulder M, Koes B, Bombardier C. Low back pain. *Best Pract Res Clin Rheumatol* 2002;16(5):761-75.
- [26] Bergquist-Ullman M, Larsson U. Acute low-back pain in industry. *Acta Orthop Scand* 1977;170:1-117.
- [27] Von Korff M, Deyo RA, Cherkin DC. Back pain in primary care: outcomes at 1 year. *Spine* 1993;18:855-62.
- [28] González Viejo MA, Condon Huerta MJ. Incapacidad por dolor lumbar en España. *Med Clin (Barc)* 2000;114:491-2.
- [29] Andersson G. Epidemiology of spinal disorders. In: Frymoyer JW, ed. *The Adult Spine: Principles and Practice*. New York: Raven Press; 1991.
- [30] Deyo RA, Mirza SK, Martin BI. Back pain prevalence and visit rates: estimates from U.S. national surveys, 2002. *Spine* 2006;31(23):2724-7.
- [31] Macfarlane GJ, Jones GT, Hannaford PC. Managing low back pain presenting to primary care: where do we go from here? *Pain* 2006;122(3):219-22.

- [32] Gureje O, von Korff M, Simon GE, Gater R. Persistent pain and well-being: A World Health Organization study in primary care. *JAMA* 1998; 280:147-51.
- [33] Dagenais S, Caro J, Haldeman S. A systematic review of low back pain cost of illness studies in the United States and internationally. *Spine J* 2008;8(1):8-20.
- [34] Druss BG, Rosenheck RA, Sledge WH. Health and Disability Costs of Depressive Illness in a Major U.S. Corporation. *Am J Psychiatry* 2000;157:1274-8.
- [35] Freburger JK, Holmes GM, Agans RP, Jackman AM, Darter JD, Wallace AS, Castel LD, Kalsbeek WD, Carey TS. The rising prevalence of chronic low back pain. *Arch Intern Med* 2009;169:251-8.
- [36] Compton WM, Conway KP, Stinson FS, Grant BF. Changes in the prevalence of major depression and comorbid substance use disorders in the United States between 1991–1992 and 2001–2002. *Am J Psychiatry* 2006; 163:2141-7.
- [37] Deyo RA, Cherkin D, Conrad D, Volinn E. Cost, controversy, crisis: low back pain and the health of the public. *Annu Rev Public Health* 1991;12:141-56.
- [38] Van Tulder MW, Koes BW, Bouter LM. A cost-illness study of back pain in the Netherlands. *Pain* 1995;62:233-40.
- [39] Van Tulder M. Chapter 1: Introduction. *Eur Spine J* 2006;15(suppl 2):S134-S135.
- [40] Norlund AI, Waddell G. Cost of back pain in some OECD countries. In: Nachemson AL, Jonsson E (eds). *Neck and back pain: The scientific evidence of causes, diagnosis and treatment*. Philadelphia: Lippincott, Williams & Wilkins, 2000.
- [41] Nachemson AL, Jonsson E. *Neck and back pain. The scientific evidence of causes, diagnosis, and treatment*. Philadelphia; London: Lippincott Williams & Wilkins, 2000.

- [42] Birkmeyer NJO, Weinstein JN. Medical versus surgical treatment for low back pain: Evidence and clinical practice. *Effective Clin Pract* 1999;2(5):218-27.
- [43] Review: evidence for the effectiveness of surgery for low back pain, radiculopathy, and spinal stenosis is limited. *Evid Based Med* 2009;14:18.
- [44] Chou R, Baisden J, Carragee EJ, Resnick DK, Shaffer WO, Loeser JD. Surgery for low back pain: a review of the evidence for an American Pain Society Clinical Practice Guideline. *Spine* 2009;34(10):1094-109.
- [45] Mirkovic S, Garfin SR, Rydevik B, Lipson SJ. Pathophysiology of spinal stenosis. *Instr Course Lect* 1992;41:165-77.
- [46] Herkowitz HN. Spinal stenosis: clinical evaluation. *Instr Course Lect* 1992;41:183-5.
- [47] Deyo RA, Gray DT, Kreuter W, et al. United States trends in lumbar fusion surgery for degenerative conditions. *Spine* 2005;30:1441-5.
- [48] Schonstrom NSR, Bolender NF, Spengler DM. The pathomorphology of spinal stenosis as seen on CT scans of the lumbar spine. *Spine* 1985;10:806-11.
- [49] Boden SD, McCowin PR, Davis DO, et al. Abnormal magnetic-resonance scans of the lumbar spine in asymptomatic subjects: a prospective investigation. *J Bone Joint Surg Am* 1990;72:403-8.
- [50] Atlas SJ, Delitto A. Spinal stenosis. Surgical versus non-surgical treatment. *Clin Orthop and Related Res* 2006;443:198-207.
- [51] North American Spine Society. Clinical Guidelines for Multidisciplinary Spine Care. Diagnosis and Treatment of Degenerative Lumbar Spinal Stenosis, 2007.
- [52] Deyo RA. Back surgery – Who needs it? *N Engl J Med* 2007;356(22):2239-43.

- [53] Carragee EJ, Ghanayem AJ, Weiner BK, Rothman DJ, Bono CM. A challenge to integrity in spine publications: years of living dangerously with the promotion of bone growth factors. *Spine J* 2011;11(6):463-8.
- [54] Carragee EJ, Hurwitz EL, Weiner BK. A critical review of recombinant human bone morphogenetic protein-2 trials in spinal surgery: emerging safety concerns and lessons learned. *Spine J* 2011;11(6):471-91.
- [55] Kovacs FM, Olabe J, Urrútia G, Zamora J, Bagó J, Alvarez-Galovich L, Abreira V. Yet another reason for improving approval and surveillance processes for health technologies. *Spine J*. 2011 Jul 19. [Epub ahead of print]
- [56] Evidence-Based Medicine Working Group. Evidence-based medicine. A new approach to teaching the practice of medicine. *JAMA* 1992;268(17):2420-5.
- [57] Deyo RA. Gaps, tensions, and conflicts in the FDA approval process: implications for clinical practice. *J Am Board Fam Pract* 2004;17(2):142-9.
- [58] Institute of Medicine. *Clinical Practice Guidelines We Can Trust*. Washington, DC: The National Academies Press, 2011.
- [59] McLeod RS, Wright JG, Solomon MJ, Hu X, Waiters BC, Lossing A. Randomized controlled trials in surgery. Issues and problems. *Surgery* 1996;119:483-6.
- [60] Ergina PL, Cook JA, Blazeby JM, Boutron I, Clavien PA, Reeves BC, Seiler CM, for the Balliol Collaboration. Challenges in evaluating surgical innovation. *Lancet* 2009;374:1097-104.
- [61] Furlan AD, Pennick V, Bombardier C, van Tulder M; Editorial Board, Cochrane Back Review Group. 2009 updated method guidelines for systematic reviews in the Cochrane Back Review Group. *Spine* 2009;34(18):1929-41.
- [62] Corcoll J, Orfila J, Tobajas P, Alegre L. Implementation of neuroreflexotherapy for subacute and chronic neck and back pain within the Spanish public health system: audit results after one year. *Health Policy* 2006;79:345-57.

- [63] Amundsen T, Weber H, Nordal HJ, Magnaes B, Abdelmoor M, Lilleas F. Lumbar spinal stenosis: conservative or surgical management? A prospective 10-year study. *Spine* 2000;25:1424-36.
- [64] Weinstein JN, Lurie JD, Tosteson TD, et al. Surgical versus nonsurgical treatment for lumbar degenerative spondylolisthesis. *N Engl J Med* 2007;356(22):2257-70.
- [65] Weinstein JN, Tosteson TD, Lurie JD, et al. An H. Surgical versus nonsurgical therapy for lumbar spinal stenosis. *N Engl J Med* 2008;358:794-810.
- [66] Gibson JN, Waddell G. Surgery for degenerative lumbar spondylosis. *Cochrane Database Syst Rev* 2005 Oct 19;(4):CD001352.
- [67] <http://www.kovacs.org/reide.asp> [consultat el 10 de juny de 2011].
- [68] Freeman-Brian JC. IDET: a critical appraisal of the evidence. *Eur Spine J* 2006;15(Suppl 3):S448-S457.
- [69] <http://journals.lww.com/spinejournal/blog/SpineBlog/pages/default.aspx>. [consultat el 19 de setembre de 2011].
- [70] Schulz KF, Altman DG, Moher D, for the CONSORT Group CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;340:c332.
- [71] Boutron I, Moher D, Altman DG, Schulz K, Ravaud P, for the CONSORT group. Extending the CONSORT Statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. *Ann Intern Med* 2008;148:295-309.
- [72] Boutron I, Tubach F, Giraudeau B, Ravaud P. Methodological differences in clinical trials evaluating nonpharmacological and pharmacological treatments of hip and knee osteoarthritis. *JAMA* 2003;290:1062-70.

- [73] Jacquier I, Boutron I, Moher D, Roy C, Ravaud P. The reporting of randomized clinical trials using a surgical intervention is in need of immediate improvement: a systematic review. *Ann Surg* 2006;244:677-83.
- [74] Stephenson J, Imrie J. Why do we need randomised controlled trials to assess behavioural interventions? *BMJ* 1998;316:611-3.
- [75] Tiruvoipati R, Balasubramanian SP, Atturu G, Peek GJ, Elbourne D. Improving the quality of reporting randomized controlled trials in cardiothoracic surgery: the way forward. *J Thorac Cardiovasc Surg.* 2006;132(2):233-40.
- [76] Balasubramanian SP, Wiener M, Alshameeri Z, Tiruvoipati R, Elbourne D, Reed MW. Standards of reporting of randomized controlled trials in general surgery: can we do better? *Ann Surg* 2006;244(5):663-7.
- [77] Sinha S, Sinha S, Ashby E, Jayaram R, Grocott MP. Quality of reporting in randomized trials published in high-quality surgical journals. *J Am Coll Surg* 2009;209(5):565-71.
- [78] Agha R, Cooper D, Muir G. The reporting quality of randomised controlled trials in surgery: a systematic review. *Int J Surg* 2007;5(6):413-22.
- [79] Richards D. Quality of reporting randomised controlled trials in dental public health. *Evid Based Dent* 2011;12(2):54.
- [80] Wang G, Mao B, Xiong ZY, Fan T, Chen XD, Wang L, Liu GJ, et al. The quality of reporting of randomized controlled trials of traditional Chinese medicine: a survey of 13 randomly selected journals from mainland China. *Clin Ther* 2007;29(7):1456-67.
- [81] Linde K, Jonas WB, Melchart M, Willich S. The methodological quality of randomized controlled trials of homeopathy, herbal medicines and acupuncture. *International Journal of Epidemiology* 2001;30:526-31.
- [82] Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P, for the CONSORT Group. Methods and Processes of the CONSORT Group: Example of an Extension for Trials Assessing Nonpharmacologic Treatments. *Ann Intern Med.* 2008;148:60-6.

- [83] Deyo RA, Battie M, Beurskens AJ, Bombardier C, Croft P, Koes B, Malmivaara A, Roland M, Von Korff M, Waddell G. Outcome measures for low back pain research. A proposal for standardized use. *Spine* 1998;23:2003-13.
- [84] Wilmshurst P. The regulation of medical devices. *BMJ* 2011;342:d2822.
- [85] Fraser AG, Krucoff MW, Brindis RG, Komajda M, Smith SC Jr. Commentary: International collaboration needed on device clinical standards. *BMJ* 2011;342:d2952.
- [86] Cohen D, Billingsley M. Europeans are left to their own devices. *BMJ* 2011;342:d2748.
- [87] Cohen D. Out of joint: the story of the ASR. *BMJ* 2011;342:d2905.
- [88] Thompson M, Heneghan C, Billingsley M, Cohen D. Medical device recalls and transparency in the UK. *BMJ* 2011;342:d2973.
- [89] Carragee EJ, Deyo RA, Kovacs FM, Peul WC, Lurie JD, Urrútia G, Corbin TP, Schoene ML. Clinical research: is the spine field a mine field? *Spine* 2009;34(5):423-30.
- [90] Deyo RA, Mirza SK. Trends and variations in the use of spine surgery. *Clin Orthop Relat Res* 2006;443:139-46.
- [91] Weinstein JN, Lurie JD, Olson PR, Bronner KK, Fisher ES. United States' trends and regional variations in lumbar spine surgery: 1992-2003. *Spine* 2006;31(23):2707-14.
- [92] Variaciones en cirugía ortopédica y traumatología en el Sistema Nacional de Salud. Atlas de variaciones en la práctica médica en el sistema nacional de salud. Marzo 2005. Volumen 1. Número 1
- [93] http://www.accessdata.fda.gov/cdrh_docs/pdf/k990474.pdf [consultat el 10 de juny de 2008].
- [94] Deyo R. Gaps, Tensions, and Conflicts in the FDA Approval Process: Implications for Clinical Practice. *J Am Board Fam Pract* 2004;17:142-9.

- [95] http://www.idetprocedure.com/2000_professional/2100_about.html
[consultat el 10 de juny de 2008].
- [96] Saal JS, Saal JA. Management of chronic discogenic low back pain with a thermal intradiscal catheter. A preliminary report. *Spine* 2000;25(3):382-8.
- [97] http://global.smith-nephew.com/master/news_agree_acquire_oratec_3149.html [consultat el 10 de juny de 2011].
- [98] Pauza KJ, Howell S, Dreyfuss P, et al. A randomized, placebo-controlled trial of intradiscal electrothermal therapy for the treatment of discogenic low back pain. *Spine J* 2004;4:27-35.
- [99] Freeman BJC, Fraser RD, Cain CMJ, et al. A randomized, double-blind, controlled trial: intradiscal electrothermal therapy versus placebo for the treatment of chronic discogenic low back pain. *Spine* 2005;30:2369-77.
- [100] American Society of Interventional Pain Physicians. Interventional Techniques: Evidence-based Practice Guidelines in the Management of Chronic Spinal Pain. *Pain Physician* 2007;10(1):7-111.
- [101] <http://www.crd.york.ac.uk/CRDWeb/ShowRecord.asp?ID=32003000559>
[accedit el 15 de juliol de 2011].
- [102] <http://www.msac.gov.au> [accedit el 15 de juliol de 2011].
- [103] <http://www.sst.dk/Applikationer/cemtv/publikationer/docs/ondtiryggen3123-10-1995/danskhtml/ondtiryggen/ryg.htm> [consultat el 10 de juny de 2008].
- [104] <http://www.nice.org.uk/guidance/index.jsp?action=article&o=43964>
[consultat el 10 de juny de 2008].
- [105] <http://www.aetmis.gouv.qc.ca/site/249.376.0.0.1.0.phtml> [consultat el 10 de juny de 2008].
- [106] Blue Cross Blue Shield Association Technology Evaluation Center. Percutaneous intradiscal radiofrequency thermocoagulation for chronic

discogenic low back pain. TEC Assessment In Press; October 2003. Kaiser Permanente.

[107] Smith R. Beyond conflict of interest. Transparency is the key. *BMJ* 1998;317:291-2.

[108] Moynihan R. Key opinion leaders: independent experts or drug representatives in disguise? *BMJ* 2008;336:1402-3.

[109] Smith R (2005) Medical journals are an extension of the marketing arm of pharmaceutical companies. *PLoS Med* 2(5): e138.

[110] Angell M. The truth about drug companies: How they deceive us and what to do about it. New York: Random House; 2005. 336 pp.

[111] Huston P, Moher D. Redundancy, disaggregation, and the integrity of medical research. *Lancet* 1996;347:1024-6.

[112] Hargreaves S. Ghost authorship of industry funded drug trials is common, say researchers. *BMJ* 2007;334:223.

[113] Johnson T. Shattuck lecture-medicine and the media. *N Engl J Med* 1998;339(2):87-92

[114] Entwistle V. Reporting research in medical journals and newspapers. *BMJ* 1995;310(6984):920-3.

[115] Woloshin S, Schwartz LM. Press releases: translating research into news. *JAMA* 2002;287(21):2856-8.

[116] Schwartz LM, Woloshin S, Baczek L. Media coverage of scientific meetings: too much, too soon? *JAMA* 2002;287(21):2859-63.

[117] Moynihan R, Bero L, Ross-Degnan D, Henry D, Lee K, Watkins J, Mah C, Soumerai SB. Coverage by the news media of the benefits and risks of medications. *N Engl J Med* 2000;342(22):1645-50.

[118] Gelberman RH, Samson D, Mirza SK, Callaghan JJ, Pellegrini VD Jr. Orthopaedic surgeons and the medical device industry: the threat to scientific integrity and the public trust. *J Bone Joint Surg Am* 2010;92(3):765-77.

- [119] Brand RA, Buckwalter JA, Talman CL, Happe DG. Industrial support of orthopaedic research in the academic setting. *Clin Orthop Relat Res* 2003;412:45-53.
- [120] Camilleri M, Parke DW 2nd. Perspective: Conflict of interest and professional organizations: considerations and recommendations. *Acad Med* 2010;85(1):85-91.
- [121] Kesselheim AS, Mello MM, Studdert DM. Strategies and practices in off-label marketing of pharmaceuticals: a retrospective analysis of whistleblower complaints. *PLoS Med* 2011;8(4):e1000431.
- [122] Moreno J, Gestoso M, Kovacs FM. La efectividad de la intervención neurorreflejo terapéutica en el tratamiento de la patología mecánica crónica del raquis : resultados preliminares. *Medicina del Trabajo* 1992;1:433-43.
- [123] Kovacs FM, Abaira V, López-Abente G, Pozo F. La intervención neurorreflejo terapéutica en el tratamiento de la lumbalgia inespecífica: un ensayo clínico controlado, aleatorizado, a doble ciego. *Med Clin (Barc)* 1993;101(15):570-5.
- [124] Kovacs FM, Abaira V, Pozo F, Kleinbaum DG, Beltrán J, Mateo I, Pérez de Ayala C, Peña A, Zea A, González-Lanza M, Morillas L. Local and remote sustained trigger point therapy for exacerbations of chronic low back pain. A randomized, double-blind, controlled, multicenter trial. *Spine* 1997;22(7):786-97.
- [125] Kovacs FM, Llobera J, Abaira V, Lázaro P, Pozo F, Kleinbaum D; KAP Group. Effectiveness and cost-effectiveness analysis of neuroreflexotherapy for subacute and chronic low back pain in routine general practice: a cluster randomized, controlled trial. *Spine* 2002;27(11):1149-59.
- [126] Urrútia G, Burton A, Morral A, Bonfill X, Zanolli G. Neuroreflexotherapy for non-specific low-back pain. *Cochrane Database Syst Rev* 2004;2:CD003009.
- [127] Conde JL. Informe sobre neurorreflejo terapia en el tratamiento de la lumbalgia inespecífica realizado a solicitud del Ministerio de Sanidad y Consumo. Agencia de Evaluación de Tecnologías Sanitarias. Madrid: Ministerio de Sanidad y Consumo, 20 de febrero de 1996.

[128] Sampietro L, Pons JMV. La neuroreflexoteràpia en el tractament del dolor lumbar inespecífic. Barcelona: Agència d'Avaluació de Tecnologia i Recerca Mèdiques, 2001.

[129] García Caeiro AL, Martínez Calvo AV. Informe sobre neuroreflejo terapia en el tratamiento de la lumbalgia inespecífica. Santiago de Compostela: Axencia de Avaliación de Tecnoloxías Sanitarias de Galicia (avalia-t); noviembre 2002.

[130] Porta M. La intervención neuroreflejo terapica en el Sistema Nacional de Salud español. Barcelona: Sociedad Española de Salud Pública y Administración Sanitaria (SESPAS), 31 de enero de 2003.

[131] Subdirección General de Asistencia Sanitaria. Informe sobre la propuesta de implantación mediante uso tutelado en el SNS de la intervención neuroreflejo terapica (NRT) para el tratamiento de la lumbalgia inespecífica. Madrid: Dirección General del INSALUD, 29 de octubre de 2001.

[132] Gerencia de Atención Primaria. Protocolo de derivación a intervención NRT de los pacientes con dolencias comunes de espalda. Palma de Mallorca: Dirección Territorial del INSALUD en Baleares, 4 de mayo de 2001.

[133] Subdirección General de Concursos. Especificaciones técnicas: requisitos mínimos. Intervenciones neuroreflejo terapicas. Madrid: Dirección General del INSALUD, 3 de diciembre de 2001.

[134] Direcció General d'Avaluació i Acreditació. La transferencia de la realización de intervenciones neuroreflejo terapicas al Ib-Salut. Palma de Mallorca: Conselleria de Salut i Consum del Govern de les Illes Balears, 2002.

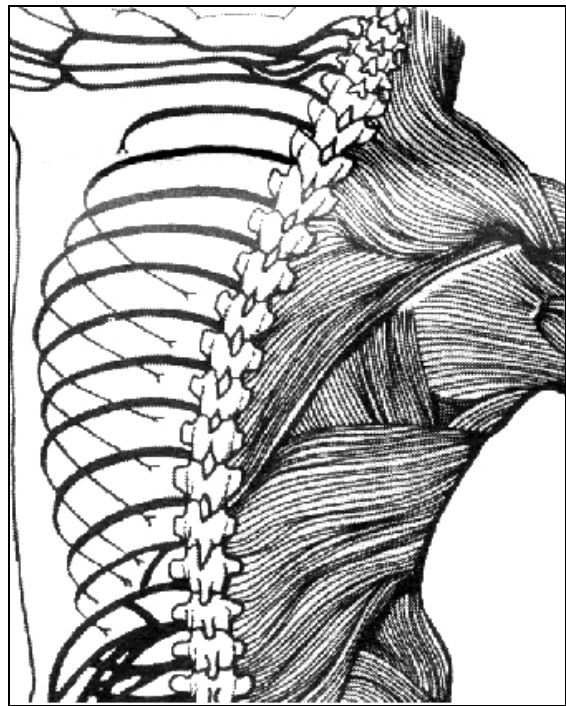
[135] Grupo PINS. Intervención neuroreflejo terapica para el tratamiento de las enfermedades mecánicas del raquis. Resultados de una experiencia piloto. Gac Sanit 2004;18(4):275-86.

[136] Kovacs F, Abreira V, Muriel A, Corcoll J, Alegre L, Tomas M, Mir MA, Tobajas P, Gestoso M, Mufraggi N, Gil del Real MT, Zamora J. Prognostic factors for neuroreflexotherapy in the treatment of subacute and chronic neck

and back pain: a study of predictors of clinical outcome in routine practice of the Spanish National Health Service. *Spine* 2007;32(15):1621-8.

[137] Urrútia G, Bonfill X. [PRISMA declaration: a proposal to improve the publication of systematic reviews and meta-analyses]. *Med Clin (Barc)* 2010;135(11):507-11.

9. Annexos



9.- ANNEXOS

Annex 1: *Urrútia G, Kovacs FM, Seco J. The efficacy, safety and effectiveness of laser therapy for low back pain: A systematic review. [Enviat a publicació]*

Annex 2: *Ferreira I, Urrútia G, Alonso-Coello P. Systematic Reviews and Meta-Analysis: Scientific Rationale and Interpretation. Rev Esp Cardiol. 2011;64(8):688–696.*

Annex 3: *Urrútia G, Bonfill X. Cap. 15. Revisiones sistemáticas. En: Publicación científica biomédica. Cómo escribir y publicar un artículo de investigación. J. Jiménez Villa, J.M. Argimon Pallàs, A. Martín Zurro, M. Vilardell Tarrés. Elsevier Españ, Barcelona (2010), ISBN: 978-84-8086-461-9*

Annex 4: *Urrútia G, Bonfill X. Declaración PRISMA: una propuesta para mejorar la publicación de revisiones sistemáticas y metanàlisis. Med Clin (Barc). 2010 Oct 9;135(11):507-11.*

Annex 5: *Carragee EJ, Deyo RA, Kovacs FM, Peul WC, Lurie JD, Urrútia G, Corbin TP, Schoene ML. Clinical research: is the spine field a mine field? Spine (Phila Pa 1976). 2009 Mar 1;34(5):423-30.*

The efficacy, safety and effectiveness of low power laser therapy for low back pain: A systematic review.

Gerard Urrutia, MD,^{1,2} Francisco M. Kovacs, MD, PhD,^{3,2} Jesús Seco, MD, PhD,^{4,2}

¹ Iberoamerican Cochrane Center-Servei d'Epidemiologia Clínica i Salut Pública, Institut d'Investigació Biomèdica Sant Pau, Barcelona, Spain and CIBER Epidemiología y Salud Pública (CIBERESP), Spain.

² Spanish Back Pain Resesarch Network.

³ Fundació Kovacs, Scientific Department, Palma de Mallorca, Spain.

⁴ Department of Nursing and Physical Therapy. Institute of Biomedicine (IBIOMED). University of León, Ponferrada, Spain.

Number of pages (including manuscript, tables and figures): 47

Number of tables:5,

Number of figures: 3

Corresponding author: Francisco M. Kovacs. Departamento Científico. Fundación Kovacs. Paseo Mallorca 36. 07012 Palma de Mallorca. Spain. Tel: +34 971 720809. Fax: +34 971 720774. E-mail: kovacs@kovacs.org

All the authors contributed substantially to the conception and design of this study, and to the revision of this article, take public responsibility for the appropriateness of the its design and method, and the collection, analysis and interpretation of the data. All the authors have reviewed and approved the final version of this manuscript for submission and/or publication.

Gerard Urrútia, Francisco Kovacs and Jesús Seco retrieved the references identified through the electronic search and assessed their suitability for inclusion in this review. Gerard Urrútia, Francisco Kovacs and Jesús Seco independently assessed the methodological quality of the studies which were included. Francisco Kovacs and Gerard Urrutia resolved disagreements in quality assessment and data extraction, by consensus.

Francisco M. Kovacs wrote this manuscript, had full access to all of the data in the study, and takes responsibility for the integrity of the data and the accuracy of the data analysis, as well as the decision to submit it for publication.

Abstract

Low Power Laser Therapy (LPLT) is commonly used for treating low back pain (LBP). The aim of this systematic review was to assess its efficacy, safety, effectiveness and cost/effectiveness. An electronic search was conducted in MEDLINE, EMBASE and the Cochrane Library databases up to July 2009, to identify randomised controlled trials (RCTs) comparing LPLT to placebo or to other treatments for LBP. No language restrictions were applied. Additional data were requested from the authors of the original studies. The risk of bias of each study was assessed following the criteria recommended by the Cochrane Back Review Group. Ten randomised controlled trials (RCTs) were included. Nine had an acceptable internal validity, five gathered adverse events, and only two disclosed sources of funding. None assessed cost/effectiveness. LPLT showed to be better than an inert sham procedure for treating an LBP episode and possibly for reducing relapses within 3-12 months. For common LBP, results from LPLT, exercise and the combination of LPLT and exercise were similar. For pain caused by disc herniation, the addition of LPLT to NSAID improved patients' evolution at 3 weeks, whereas a sub-optimal dose of LPLT was similar to traction and ultrasound. No data on the clinical relevance of the effects triggered by LPLT, were identified. LPLT is superior to an inert sham procedure for treating LBP. Evidence on clinical relevance and cost/effectiveness would be necessary to recommend the generalisation of this procedure in routine clinical practice.

Key words: low back pain, laser, treatment, effectiveness, randomised controlled trials, systematic review.

Introduction

Nonspecific or common low back pain (LBP) is defined as pain between the costal margins and the inferior gluteal folds, which is usually accompanied by painful limitation of movement and may be associated with pain referred down to the leg (“leg pain”) [50]. Diagnosing common LBP implies that the pain is not related to conditions such as fractures, spondylitis, direct trauma, or neoplastic, infectious, vascular, metabolic, or endocrine-related processes [50]. LBP, and especially chronic LBP, represents a major health care, social and economic burden [3,32,35,48,50].

Low power laser therapy (LPLT) consists in applying light energy with a specific wavelength at a predetermined energy density, which penetrates tissue and is absorbed by the cells. In 2003, the North American Association of Laser Therapy (NAALT) adopted the term phototherapy, defining it as the use of photons from the visible and infra-red spectrum to stimulate tissue repair and pain relief [34]. The most frequently used types of LPLT are the gallium-arsenide and the helium-neon lasers, which emit radiation on a wavelength of between 600 and 984 nm [5,7,22,24].

LPLT generates localised heat. It has been claimed that it may also trigger photochemical reactions in the cells, which would stimulate or inhibit biochemical, physiological and proliferative cell activity [4,7], stimulate cell growth, increase cell metabolism, accelerate repair of connective tissue, regulate fibroblast function, reduce oedema, inhibit action potentials in nerve roots, reduce production of substance P, stimulate long-term production of nitric oxide, inhibit the release of histamine and acetylcholine, reduce synthesis of prostaglandins, and stimulate the production of endorphins [29,36,40,46]. As a result, it has been suggested that LPLT might reduce pain and inflammation, and stimulate regeneration of the neural, muscular and bony tissue [16]. However, to date most mechanisms explaining these potential effects remain hypothetical [4,16].

LPLT is used to treat musculoskeletal disorders, including LBP. Approximately 50% of UK physiotherapists, 65% of US physiotherapists, and 94% of Canadian physiotherapists, use it [30], and approximately 55% of US primary care practitioners recommend it [18], despite the fact that the most recent Cochrane systematic review concludes that there is insufficient evidence to draw firm conclusions on its effectiveness [52], and that the most recent evidence based guidelines for the treatment of LBP do not recommend it [1,14,49]. However, additional randomised controlled trials (RCTs) have been published after the period covered by the Cochrane review, and a recent meta-analysis suggests that LPLT is effective for neck pain [15].

Therefore, the objective of this study was to systematically review the evidence on the efficacy, effectiveness, safety, and cost/effectiveness of LPLT for treating LBP.

Materials and methods

The protocol of this study was approved by and registered in the (*information masked for confidentiality during reviewing process*) and the (*information masked for confidentiality during reviewing process*) databases.

Search and Study Selection

A comprehensive literature search was conducted to identify relevant studies using the following bibliographic databases for the periods indicated: i) CENTRAL (January 2010), ii) MEDLINE (1970 to January 2010), iii) EMBASE (1989 to March 2009), iv) CINAHL (1987 to July 2009), v) AMED (Allied and Complementary Medicine Database) (1987 to March 2009) and vi) PEDro (Physiotherapy Evidence Database, <http://www.pedro.fhs.usyd.edu.au/index.html>) (to March 2009). The search strategy was designed to ensure maximum sensitivity and is shown in Table 1. No language restrictions were applied.

Studies were included in this review if they: a) Were published randomised controlled clinical trials (RCT) which assessed the effects of any type of LPLT to treat LBP, and compared them to sham intervention, placebo or any other active treatment, b) Assessed at least one of the outcome variables considered to be relevant: pain (LBP or pain irradiated down to the leg; leg pain –LP-), disability, health-related quality of life, daily activity, work-related variables (e.g., absenteeism or return to work), patients' satisfaction, use of drugs, adverse events, use of health resources or costs [17].

References identified through the electronic search were screened. The full text of those which were eligible was assessed for inclusion criteria separately by two authors (*information masked for confidentiality during the reviewing process*). Disagreements were resolved by consensus with a third author (*information masked for confidentiality during the reviewing process*). References of potential additional studies were searched for in the reference sections of the studies included.

Risk of Bias Assessment and Data Analysis

Following the recommendations of the Cochrane Back Review Group, the risk of bias of the studies included was assessed separately by two of the reviewers (*information masked for confidentiality during the reviewing process*), and disagreements were solved by consensus with the third author (*information masked for confidentiality during the reviewing process*) [21]. Following these recommendations, a study was rated as having a “low risk of bias” when it met ≥ 6 of 11 methodological criteria. For studies which compared procedures which did not allow masking of therapists and patients, these two criteria were scored as “not applicable”. For the latter, the maximum possible score was 9, and they were considered as “low risk of bias” if they scored ≥ 5 . However, studies with serious flaws were considered as “high risk of bias”, regardless of their score [21]. When a study lacked information required to assess the risk of bias, missing data were requested directly from its authors.

Data extraction was undertaken separately and in duplicate, using standardized forms. For each study, data on pain severity, disability, quality of life, daily activities, impact on work, patient satisfaction, medication use, side effects, health resources employed, and total costs, were searched for, and all data on all the variables gathered in each individual study were extracted.

The LPLT treatment protocols used in studies included in this review, were assessed to determine whether they complied with the standards issued by the World Association for Laser Therapy (WALT) in 2010, on the minimal “dose” required to be considered as LPLT [51]. These standards are as follows: treatment daily for 2 weeks, or every other day for 3-4 weeks, with therapeutic doses ranging \pm 50% of the following: For GaAlAs 780 - 860nm laser: continuous or pulsed mode with a mean output <0.5 Watt, applied to 4-8 points or cm^2 for 20 to 300 seconds, with a total dose of 16 Joules and a minimum of 4 Joules per point. For GaAs 904 nm laser: pulses (peak pulse output >1 W, mean output >5 mW and power density $>5\text{mW}/\text{cm}^2$), applied to 4 points or cm^2 with a total dose of at least 4 Jules and a minimum of 1 Joules per point. According to the WALT, doses outside these windows are inappropriate and should not be considered as LPLT [51].

A quantitative synthesis of data (metanalysis) was not possible because of the variability of outcome measures and heterogeneity of methods across studies. Therefore, a qualitative analysis of data was performed, based on the methodological quality of the trials included and the consistency of their findings.

In addition to the analysis in which all studies were included, a sensitivity analysis was performed. In the latter, only studies providing a laser treatment protocol complying with the WALT definition of LPLT, were included.

Role of the funding sources

The funding institutions had no role in the design and conduction of the study; data collection, management, analysis and interpretation; preparation, review and approval of the manuscript; or in the decision to submit the article for publication.

Results

The electronic search provided 1,419 references. Once duplicate references had been eliminated, 134 potentially relevant studies were identified and 13 were initially selected for this review [6,19,23,25,26,27,31,33,39,43,44,45,47]. Three out of these 13 were excluded; one because it had included patients with both neck pain and LBP without providing separate results for those with LBP [43], a second one because it compared a non-steroidal anti-inflammatory drug (NSAID) with a combination of LPLT, TENS and exercise –making it impossible to assess the effect of LPLT- [26], and a third one because it was the summary of a conference which only presented results in a narrative form [39].

The remaining ten studies had recruited a total of 1053 patients, and were included in this systematic review [6,19,23,25,27,31,33,44,45,47]. Their characteristics are shown in Table 2, and their methodological assessment is shown in Table 3. There was no disagreement between reviewers with regards to data extraction and the assessment of risk of bias.

Across the 10 studies included in this review, the method of randomization was adequate in five [6,19,25,27,45], inadequate in one [33], and insufficiently described in four [23,31,44,47] (Tables 2 and 3).

Among the ten studies included in this review, five used a laser device equipped with a 904 nm gallium arsenide diode (GAS) [23,25,27,31,44], three a 780-860 nm Gallium-

Aluminum Arsenide (GaAlAs) diode [19,45,47], one a He-Ne laser [33], and the remaining one a Nd-YAG laser, which has some thermal effect [6].

The laser was applied in continuous mode in four studies [6,19,45,47], and in pulse mode in five studies [23,25,27,31,44]. One study did not specify whether the laser was used in pulse or continuous mode [33].

Treatment regimes varied across studies; one study applied LPLT only once [45], another one applied 2 sessions per week for 6 weeks [19]; two applied 3 sessions a week for 4 weeks [6,25], two applied 4 sessions a week for 3 weeks [27,47] (Table 2). One applied one session per day for 5 days, followed by 5 sessions on alternative days [31], and the remaining three applied 5 sessions a week for 2 weeks [23,33,44]. Seven studies [6,19,23,25,27,31,44] used a laser dosage complying with the WALT recommendations issued in 2010 [51].

Two studies did not provide any information on cointerventions [23,33]. Other studies report that the patients “did not receive other therapies” [44,45], that “exercises were prescribed to all patients in both groups” [25], or that “patients were asked to take as few analgesics as possible and a record was kept of analgesics taken” [6]. (Table 2).

Across these studies LPLT was compared to a sham inactive procedure, defined as the application of the same device without emission of laser radiation [6,19,25,27,31,44,45], exercise [23], exercise+laser [19,23], NSAID [27], ultrasound [47], traction [47], and a program combining mesotherapy with non-steroidal antiinflammatory drugs (NSAIDs), postural education, exercise and a lumbar support [33].

Four studies included only acute patients [27,31,33,47], four only chronic patients [19,23,25,44], one both acute and chronic patients [6], and one did not disclose patients’ pain duration [45]. Patients with pain radiating down into the leg were accepted in four studies [19,27,45,47], two of which recruited only patients with radicular pain caused by a disc herniation confirmed by MRI (Table 2) [27,47].

All the studies assessed the evolution of pain severity [6,19,23,25,27,31,33,44,45,47]. Among these, six used previously validated instruments and also measured disability and range of motion [6,19,23,25,27,47]. Other outcomes were pain relapses [31,44], quality of life [27], Men’s stability test [33], and straight leg raising test [47]. Adverse events were gathered in five studies [6,19,27,44,45].

Patients, care providers and outcome assessors were blinded in seven studies [6,19,25,27,31,44,45]. In one study, only the outcome assessor was blinded [23]. In the remaining two studies, the nature of interventions compared, made it impossible to blind either patients or care providers; in one of these studies the outcome assessor was blinded [47]. The latter was not disclosed in the remaining study [33]. (Tables 2 and 3).

Follow-up period ranged between 1 day and 1 year (Table 2) [31,33,45]. In one study, drop-out rate was not described [33], in another one it was high [44], and in the remaining eight it was acceptable (Table 3). Analyses were by intention to treat in six studies [19,23,25,27,45,47]. This was not the case in three studies [6,31,44], and it is unclear in the remaining one [33].

One study was funded by a US research foundation [25], and another one by a manufacturer of laser equipment (LaserBiotherapy Inc) [6]. The source of funding was not disclosed in the remaining eight studies [19,23,27,31,33,44,45,47].

Six studies compared LPLT to a sham inert procedure [6,19,25,27,31,45], although two of them added a cointervention in each group; a non-steroidal anti-inflammatory drug (NSAID) in one study [27], and exercise in the other one [19]. Among these six studies, four assessed pain, disability and range of motion with validated instruments [6,19,25,27].

Results from the studies included in this review are shown in Table 4, whereas they are shown separately for those studies complying and not complying with WALT recommendations in Table 5 [51]. Figures 1, 2 and 3 show the forest plots corresponding to the results on pain and disability.

Among the six randomised trials assessing the effectiveness of LPLT versus a sham inert procedure, the one rated with the highest possible methodological score failed to find any differences [25], whereas LPLT was superior in the other five [6,19,27,44,45]. (Figures 1 and 2).

In two studies, results on pain, disability and range of motion obtained from LPLT and from the combination of LPLT and exercise, were similar [19,23]. Only one of these studies compared LPLT to exercise, and failed to find any differences [23].

According to a study with a high risk of bias, for patients with acute LBP and positive sacroiliac signs, LPLT was less effective than a program combining exercise, magnetotherapy, and a lumbar support [33].

For patients with radicular pain caused by disc herniation, the effectiveness of a dose of LPLT below the WALT recommendations, ultrasound and traction, were similar [47], whereas the addition of LPLT to NSAID led to a better evolution of pain, disability and range of motion than using only NSAID [27].

Discussion

Results from the studies included in this review suggest that LPLT is better than an inert sham procedure for the treatment of an episode of common LBP [6,44,45], and possibly also for reducing relapses at 6-12 months [31,44]. These results are generally consistent with those from the most recent systematic review on the effect of LPLT for neck pain [15], and with those on its effect for other musculoskeletal complaints [9,11].

All the studies included in this review which focused on such a comparison had an acceptable internal validity (Table 3), and all but one showed the superiority of actual LPLT. The exception was a study which included only 20 patients [25], which may have led to a high risk of beta error. Moreover, baseline pain severity was low (approximately 3 VAS points), (Table 3) making it difficult for differences in the evolution across groups to reach statistical significance. In addition, although the authors stated that they used a dose of 1.3 J per cm², the data they provided (laser mode: pulsed, 1 hKz pulse frequency, 200 nsec pulse duration, output power 2W, exposure time: 240 sec for each of the 5 points) [25], correspond to an actual dose of 0.1 J per cm², which is much lower than the minimal dose required for LPLT [51].

One of the studies which found LPLT to be superior to sham also used a dose below those standards, because it only applied one single session of LPLT [45]. However, the dose applied in that session was above those standards and the study focused on the effect of one single dose within one day, which may account for its results being positive.

Results from this review do not support firm conclusions on the effectiveness of LPLT compared to other forms of treatment, or on whether its addition to other forms of treatment improves results. For patients with common LBP, results from LPLT appear to be similar to those from exercise [23], although the addition of LPLT to exercise does not improve results [19,23]. For patients with LBP and radicular pain caused by a disc herniation, one study suggests that the addition of LPLT (either actual or sham) improves results from NSAID, with actual LPLT being superior to sham LPLT [27], whereas another one shows that a below-standard dose of LPLT leads to results which are similar to those from treatments which have not shown to be effective [47], such as ultrasound or traction [13,41].

The fact that there is evidence on the superiority of LPLT over a sham procedure, but not on whether adding it to other forms of treatment actually improves results, casts doubts on the clinical relevance of its effect. Only one study assessed the effect size through Cohen's δ statistic, suggesting a large effect for both actual and sham LPLT (Table 4) [27]. However, this result reflects the statistical meaningfulness of the effect, and not its clinical relevance. In fact, most differences in clinical evolution across studies included in this review are so small that they are likely to be clinically irrelevant according to the available evidence on the minimal clinically important change for pain and disability in LBP patients (Table 4) [28,38]. Therefore, further studies should assess the clinical relevance of the effect of LPLT, for instance by estimating the number needed to treat, for one patient to experience a clinically relevant improvement.

No studies assessed cost/effectiveness of LPLT, and only five, in which a total of 796 patients were included and 778 were analyzed, gathered data on adverse events [6,19,27,44,45]. Among them, four had a sample size ranging between 41 and 85 subjects [6,19,44,45], and one of them followed-up patients for only one day [45]. These features make it impossible for these studies to rule out potentially serious adverse events. In one of these studies, warmth in the treated area was spontaneously reported more frequently among patients receiving actual LPLT versus an inert sham procedure [6]. Although this does not appear to be a serious adverse event, it might have compromised patients' blindness, despite authors' claim that differences between groups were not statistically significant [6]. The highest rate of adverse events was reported in the largest study, which included 546 patients [27]. In this study, pain worsening for less than 6 hours was reported in 27 (14,8%) out of the 182 patients receiving actual LPLT in addition to NSAID, and for more than 10 consecutive days in one (0.5%) patient receiving actual LPLT and one (0.5%) receiving sham LPLT, whereas worsening of pain was not reported by any patient receiving NSAID without LPLT stimulation [27].

Application conditions may influence the effect of health technologies in clinical practice, especially for non-pharmacological technologies. For instance, the anatomical location where LPLT was applied varied across studies, and there is no evidence on whether such a location is relevant or not, or on whether location should be adapted depending on any particular clinical characteristic. Similarly, current standards on optimum dosage, wavelength, number and frequency of sessions are largely based on consensus, as opposed to solid evidence, and have changed over time [51]. Therefore, further research into those standards is needed. Training requirements for use of LPLT in routine practice should also be standardized, and be equivalent to those followed by therapists involved in the RCTs included in this review [12].

The only study funded by a research foundation led to negative results [25], whereas positive results were obtained in the only one which is known to have been funded by a manufacturer of laser equipment [6]. This is consistent with previous evidence on the correlation between funding from the industry and the reporting of positive results

[2,8,20,37,42]. Funding sources were not disclosed in the other eight studies included in this review [19,23,27,31,33,44,45,47] which makes it impossible to rule out potential influences from funding sources.

Publication bias generally favours positive results. Therefore, although efforts were made to ensure the comprehensiveness of the literature search, it is impossible to know whether unpublished studies would alter the conclusions of this review.

Given that LPLT is widely available in routine practice [18,30], that results from this review suggest that it is relatively safe, and that it appears to be better than a sham procedure, there would be no justification for recommending ceasing its clinical application. However, in order to recommend its generalisation in routine practice, potential requisites for successful application (e.g., training standards and standards on dosage, wavelength, number and frequency of sessions) should be available, and further studies should be conducted to gather data on cost/effectiveness and on the clinical relevance of its effect, especially when added to other forms of treatment which have shown to be effective.

These RCTs should ensure: a) high methodological quality, b) samples which are large enough to ensure statistical power to detect clinically meaningful differences, c) standardized treatment protocols and indication criteria, d) “intention to treat” and “as treated” analyses, e) systematic report of cointerventions and adverse events, f) assessment of costs and cost/effectiveness across treatment groups, g) quality of reporting, by following the CONSORT recommendations on reporting for trials on non-pharmacological interventions [10], including disclosure of sources of funding and potential conflicts of interest.

Registries would not substitute such studies. However, until these become available, the creation of prospective registries of patients undergoing LPLT in clinical practice would be advisable to gather further data on clinical effect and safety, and to potentially contribute to refining indication criteria, which were inconsistent across the RCTs included in this review.

In conclusion, this systematic review shows that LPLT is superior to an inert sham procedure for treating common LBP but that there is insufficient evidence on clinical usefulness and none on cost/effectiveness, which prevent recommending this procedure to be generalised in routine clinical practice.

Acknowledgements

The authors are grateful to Drs. Longo and Konstantinovic for providing further data to those included in the original report of their studies.

This study was funded by the Instituto de Salud Carlos III (grant number PI03/0809) and the Fundación Kovacs, an independent not for profit institution specialising in neck and back pain research, with no links to the health industry.

The funding institutions had no role in the design and conduct of the study; data collection; management, analysis and interpretation of the data; preparation, review and approval of the manuscript; or the decision to submit the article for publication.

Authors do not have any financial relationship with this work or to any company related to the subject of this study. They have not received and will not receive any benefits in any form from a party related directly or indirectly to the subject of this article. The authors do not have conflicts of interest to report.

References

- [1] Airaksinen O, Brox JI, Cedraschi C, Hildebrandt J, Klüber-Moffett J, Kovacs F, Mannion AF, Reis S, Staal JB, Ursin H, Zanoli G; COST B13 Working Group on Guidelines for Chronic Low Back Pain. Chapter 4. European guidelines for the management of chronic nonspecific low back pain. *Eur Spine J* 2006;15(Supplement 2);S192-300.
- [2] Als-Nielsen B, Chen W, Gluud C, Kjaergard LL. Association of funding and conclusions in randomised drug trials: a reflection of treatment effect or adverse events? *JAMA* 2003;290:921–928.
- [3] Andersson GBJ. The epidemiology of spinal disorders. In: J.W. Frymoyer, editor(s). *The adult spine: principles and practice*. Second Edition. Raven Press, Ltd, 1997:93-141.
- [4] Basford JR. The clinical and experimental status of low energy laser therapy. *Crit Rev Phys Rehabil Med* 1989; 1: 1-9.
- [5] Basford JR. Low intensity laser therapy: still not an established clinical tool. *Lasers Surg Med* 1995; 16(4): 331-342.
- [6] Basford JR, Sheffield CG, Harmsen WS. Laser therapy: a randomised, controlled trial of the effects of low-intensity Nd:YAG laser irradiation on musculoskeletal back pain. *Arch Phys Med Rehabil* 1999; 80(6): 647-652.
- [7] Baxter GD. *Therapeutic Lasers: theory and practice*. Edinburgh: Churchill Livingstone. 1994.
- [8] Bhandari M, Busse JW, Jackowski D, Montori VM, Schünemann H, Sprague S, Mears D, Schemitsch EH, Heels-Ansdell D, Devereaux PJ. Association between industry funding and statistically significant pro-industry findings in medical and surgical randomised trials. *CMAJ* 2004;170:477–480.
- [9] Bjordal JM, Couppé C, Chow RT, Tuner J, Ljunggren EA. A Systematic review of low level laser therapy with location-specific doses for pain from chronic joint disorders. *Aust J Physiother* 2003; 49(2): 107-116.
- [10] Boutron I, Moher D, Tugwell P, Giraudeau B, Poiraudeau S, Nizard R, Ravaud P. A checklist to evaluate a report of a nonpharmacological trial (CLEAR NPT) was developed using consensus. *J Clin Epidemiol* 2005;58;1233-1240.
- [11] Brosseau L, Robinson V, Wells G, Debie R, Gam A, Harman K, Morin M, Shea B, Tugwell P. (2005). Low level laser therapy (Classes I, II and III) for treating rheumatoid arthritis. *Cochrane Database of Systematic Reviews*, Issue 4. Art. No.: CD002049. DOI: 10.1002/14651858.CD002049.pub2.
- [12] Carragee EJ, Deyo RA, Kovacs FM, Peul WC, Lurie JD, Urrútia G, Corbin TP, Schoene ML. Clinical research. Is the spine field a mine field? *Spine*. 2009;34;423-430.
- [13] Clarke JA, van Tulder MW, Blomberg SE, de Vet HC, van der Heijden GJ, Bronfort G, Bouter LM. Traction for low-back pain with or without sciatica. *Cochrane Database Syst Rev*. 2007 Apr 18;(2):CD003010.
- [14] Chou R, Huffman LH; American Pain Society; American College of Physicians. Nonpharmacologic therapies for acute and chronic low back pain: a review of the evidence for an American Pain Society/American College of Physicians clinical practice guideline. *Ann Intern Med* 2007; 147(7); 492-504.
- [15] Chow RT, Johnson MI, Lopes-Martins RA, Bjordal JM. Efficacy of low-level laser therapy in the management of neck pain: a systematic review and meta-analysis of randomised placebo or active-treatment controlled trials. *Lancet*. 2009;374(9705):1897-908. Review. [Erratum in: *Lancet*. 2010; 13;375(9718):894.].
- [16] de Bie RA, Verhagen A, de Vet HCW, Lenssen T, van den Wildenberg FAJM, Kootstra A, Knipschild PG. Efficacy of 904 nm laser therapy in musculoskeletal disorders. *Phys Ther Rev* 1998; 3: 1-14.
- [17] Deyo RA, Battie M, Beurskens AJ, Bombardier C, Croft P, Koes B, Malmivaara A, Roland M, Von Korff M, Waddell G. Outcome measures for low back pain research - a proposal for standardized use. *Spine* 1998; 23(18): 2003-2013.

- [18] Di Lorio D, Henley E, Doughty A. A survey of primary care physician practice patterns and adherence to acute low back problem guidelines. *Arch Fam Med* 2000; 9; 1015-1021.
- [19] Djavid G, Mehrdad R, Ghasemi M, Hasan-Sadeh H, Sotoodeh-Manesh A, Pouryaghoub G. Low level laser therapy in chronic low back pain. *Aust J Physiother* 2007;53:155-160.
- [20] Finucane TE, Boulton CE. Association of funding and findings of pharmaceutical research at a meeting of a medical professional society. *Am J Med* 2004;117:842–845
- [21] Furlan AD, Pennick V, Bombardier C, van Tulder M, from the Editorial Board of the Cochrane Back Review Group. 2009 Updated Method Guidelines for Systematic Reviews in the Cochrane Back Review Group. *Spine* 2009;34:1929-1941.
- [22] Gerschman JA, Ruben J, Gebart-Eaglemon J. Low level laser therapy for dentinal tooth hypersensitivity. *Aust Dent J* 1994; 39(6): 353-357.
- [23] Gur A, Karakoc M, Cevik R, Nas K, Sarac AJ, Karakoc M. Efficacy of low power laser therapy and exercise on pain and functions in chronic low back pain. *Lasers Surg Med* 2003; 32(3): 233-238.
- [24] Khullar SM, Emami B, Westermark A, Haanaes HR. Effect of low-level laser treatment on neurosensory deficits subsequent to sagittal split ramus osteotomy. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 1996; 82(2): 132-138.
- [25] Klein RG, Eek BC. Low-energy laser treatment and exercise for chronic low back pain: double-blind controlled trial. *Arch Phys Med Rehabil* 1990; 71(1): 34-37.
- [26] Konstantinovic L, Devecersky G, Petronic I, Jovic S, Cutovic M, Cirovic D. Quality of life in patients with subacute low back pain treated with physiotherapy rehabilitation. *Med Pregl* 2006;59 Suppl 1:35-39.
- [27] Konstantinovic L, Kanjun Z, Milovanovic A, Cutovic M, Djurovic A, Savic V. Low-level laser therapy for acute low back pain. *Photomed Laser Surg* 2009;00:1-8.
- [28] Kovacs FM, Abaira V, Royuela A, Corcoll J, Alegre L, Cano A, Muriel A, Zamora J, del Real MT, Gestoso M, Mufraggi N. Minimal clinically important change for pain intensity and disability in patients with nonspecific low back pain. *Spine* 2007;32:2915-2920.
- [29] Kreisler M, Christoffers AB, Al-Haj H, Willershausen B, d'Hoedt B. Low level 809 nm diode laser-induced in vitro stimulation of the proliferation of human gingival fibroblasts. *Lasers Surg Med*. 2002; 30: 365-359.
- [30] Lindsay DM, Dearness J, McGinley CC. Electrotherapy usage trends in private physiotherapy practice in Alberta. *Physiother Can* 1995; 47(1); 30-34.
- [31] Longo L, Tamburini A, Monti A. Treatment with 904 nm and 10600 nm laser of acute lumbago – double blind control-laser. *Journ Eur Med Laser Ass* 1991;3:16-19.
- [32] Martin BI, Deyo RA, Mirza SK, Turner JA, Comstock BA, Hoollingworth W, Sullivan SD. Expenditures and Health Status Among Adults With Back and Neck Problems. *JAMA* 2008;299(6):656-664.
- [33] Monticone M, Barbarino A, Testi C, Arzano S, Moschi A, Negrini S. Symptomatic efficacy of stabilizing treatment versus laser therapy for sub-acute low back pain. *Eur Med Phys* 2004;40:263-268.
- [34] NAALT. North American Association of Laser Therapy. The 2003 NAALT Post Conference Standards Report. 2003. <http://www.naalt.org/post-conference.htm> [Accessed August 8th, 2011].
- [35] Nimgade A, McNeely E, Milton D, Celona J. Increased Expenditures for Other Health Conditions After an Incident of Low Back Pain. *Spine* 2010;35:769–777.
- [36] Ohno T. Pain suppressive effect of low power laser irradiation. A qualitative analysis of substance P in the rat spinal dorsal root ganglion. *J Nippon Med Sch* 1997; 64 (5): 395-400.
- [37] Okike K, Kocher MS, Mehlman CT, Bhandari M. Conflict of interest in orthopaedic research. An association between findings and funding in scientific presentations. *J Bone Joint Surg Am* 2007;89:608–613.

- [38] Ostelo RW, Deyo RA, Stratford P, Waddell G, Croft P, Von Korff M, Bouter LM, de Vet HC. Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. *Spine* 2008;33(1):90-94.
- [39] Raspopovic Z, Konstantinovic L J, Tomovic M, Mihajlovic M. Influence of applied frequencies of low level laser therapy in control of acute back pain. *Journal of the Neurological Sciences* 2001;187(Suppl 1):Abstract No. 1-46-06.
- [40] Sakurai Y, Yamaguchi M, Abiko Y. Inhibitory effect of low-level laser irradiation on LPS-stimulated prostaglandin E2 production and cyclooxygenase-2 in human gingival fibroblasts. *Eur J Oral Sci* 2000; 108: 29-34.
- [41] (AUTHORS' NAMES MASKED FOR CONFIDENTIALITY DURING THE REVIEWING PROCESS) The efficacy, safety, effectiveness, and cost/effectiveness of ultrasound and shock wave therapies for low back pain: A systematic review. *Spine J* 2011 (in press).
- [42] Shah RV, Albert TJ, Bruegel-Sanchez V, Vaccaro AR, Hilibrand AS, Grauer JN. Industry support and correlation to study outcome for papers published in *Spine*. *Spine* 2005;30:1099-1105.
- [43] Snyder-Mackler L, Barry AJ, Perkins AI, Soucek MD. Effects of Helium-Neon Laser irradiation on skin resistance and pain in patients with trigger points in the neck or back. *Phys Ther* 1989. 69(5):336-341.
- [44] Soriano F, Rios R. Gallium arsenide laser Treatment and exercise for chronic low back pain: a prospective randomised and double blind study. *Laser Therapy* 1998; 10: 175-180.
- [45] Toya S, Motegi M, Inomata K, Ohshiro T, Maeda T. Report on a computer-randomised double-blind clinical trial to determine the effectiveness of the GaAlAs (830 nm) diode laser for attenuation in selected pain groups. *Laser Therapy* 1994; 6: 143-148.
- [46] Tsuchiya K, Kawatani M, Takeshige C, Sato T, Matsumoto I. Diode laser irradiation selectively diminishes slow component of axonal volleys to dorsal roots from saphenous nerve. *Neurosci Lett* 1993; 161: 65-68.
- [47] Unlu Z, Tasci S, Tarhan S, Pabuscu Y, Islak S. Comparison of 3 physical therapy modalities for acute pain in lumbar disc herniation measured by clinical evaluation and magnetic resonance imaging. *J Manipulative Physiol Ther* 2008; 31(3); 191-198.
- [48] Van Tulder MW, Koes BW, Bouter LM. A cost-of-illness study of back pain in The Netherlands. *Pain* 1995;62:233-240.
- [49] van Tulder M, Becker A, Bekkering T, Breen A, del Real MT, Hutchinson A, Koes B, Laerum E, Malmivaara A; COST B13 Working Group on Guidelines for the Management of Acute Low Back Pain in Primary Care. Chapter 3. European guidelines for the management of acute nonspecific low back pain in primary care. *Eur Spine J*. 2006;15(Supplement 2);S169-191.
- [50] Waddell G. *The Back Pain Revolution*. London: Churchill-Livingstone, 2004:1-61.
- [51] WALT-b 2005. World association of laser therapy. Recommended anti-inflammatory dosage for Low Level Laser Therapy. <http://www.walt.nu/dosage-recommendations.html> [Accessed August, 8th, 2011].
- [52] Yousefi-Nooraie R, Schonstein E, Heidari K, Rashidian A, Pennick V, Akbari-Kamrani M, Irani S, Shakiba B, Mortaz Hejri SA, Mortaz Hejri SO, Jonaidi A. Low level laser therapy for nonspecific low-back pain. *Cochrane Database Syst Rev*. 2008 Apr 16;(2):CD005107. Review.

Figures legends:

Fig. 1: Forest plot showing the results on severity of low back pain at the end of treatment, as measured by a Visual Analog Scale

1: Djavid 2007 [19]: Differences in favour of LPLT were statistically significant at the assessment conducted at the end of the follow-up period (6 weeks after the end of treatment).

2: Comparisons were: Gur 2003 [23]: LPLT vs. Exercise vs. LPLT+exercise. Konstatinovic 2009 [27]: nimesulide versus nimesulide +LPLT. Monticone 2004 [33]: LPLT vs. a combined treatment consisting of mesotherapy, a dynamic sacroiliac support and exercises. Unlu 2008 [47]: Traction vs. Ultrasound vs. LPLT at a minimal dose

3: Gur 2003 [23] did not report the SD of mean change. There were no differences between groups.

4: Monticone 2004 [33] reported median and range, but not mean. Results tended to be favourable to the combined treatment.

Fig. 2: Forest plot showing the results on severity of low back pain at the end of treatment, as measured by patient perception of effectiveness

Fig. 3: Forest plot showing the results on disability at the end of treatment

ODI: Oswestry Disability Index. RMQ: Roland-Morris Questionnaire. MODQ: Modified Oswestry Disability Questionnaire.

1: Basford 1999 [6] reported a statistically significant difference in the mean change between groups

2: Djavid 2007 [19]: Differences in favour of LPLT were statistically significant at the assessment conducted at the end of the follow-up period (6 weeks after the end of treatment).

3: Comparisons were: Gur 2003 [23]: LPLT vs. Exercise vs. LPLT+exercise. Konstatinovic 2009 [27]: nimesulide versus nimesulide +LPLT. Monticone 2004 [33]: LPLT vs. a combined treatment consisting of mesotherapy, a dynamic sacroiliac support and exercises. Unlu 2008 [47]: Traction vs. Ultrasound vs. LPLT at a minimal dose

4: Konstantinovic 2009 [27] reports median (change from baseline) and p25-p75, instead of mean and SD. Results were favourable to the LPLT group for both comparisons ($p < 0.001$).

Focus on: Contemporary Methods in Biostatistics (III)

Systematic Reviews and Meta-Analysis: Scientific Rationale and Interpretation

Ignacio Ferreira González,^{a,b,*} Gerard Urrútia,^{b,c} and Pablo Alonso-Coello^{b,c}

^aUnidad de Epidemiología, Servicio de Cardiología, Área del Cor, Hospital Vall d'Hebron, Barcelona, Spain

^bCIBER de Epidemiología y Salud Pública (CIBERESP), Spain

^cCentro Cochrane Iberoamericano-Servei d'Epidemiologia Clínica i Salut Pública, Institut d'Investigació Biomèdica Sant Pau, Barcelona, Spain

Article history:

Available online 30 June 2011

Keywords:

Systematic review

Meta-analysis

Evidence-based medicine

ABSTRACT

Systematic reviews represent a specific type of medical research in which the units of analysis are the original primary studies. They are essential tools in synthesizing available scientific information, increasing the validity of the conclusions of primary studies, and identifying areas for future research. They are also indispensable for the practice of evidence-based medicine and the medical decision-making process. However, conducting high quality systematic reviews is not easy and they can sometimes be difficult to interpret. This special article presents the rationale for carrying out and interpreting systematic reviews and uses a hypothetical example to draw attention to key-points.

© 2011 Sociedad Española de Cardiología. Published by Elsevier España, S.L. All rights reserved.

Revisiones sistemáticas y metaanálisis: bases conceptuales e interpretación

RESUMEN

Las revisiones sistemáticas son investigaciones científicas en las cuales la unidad de análisis son los estudios originales primarios. Constituyen una herramienta esencial para sintetizar la información científica disponible, incrementar la validez de las conclusiones de estudios individuales e identificar áreas de incertidumbre donde sea necesario realizar investigación. Además, son imprescindibles para la práctica de una medicina basada en la evidencia y una herramienta fundamental en la toma de decisiones médicas. Sin embargo, la realización de una revisión sistemática de calidad no es una tarea sencilla, como en ocasiones tampoco lo es su interpretación. En este artículo especial se presentan las bases conceptuales para la realización y la interpretación de revisiones sistemáticas, poniendo especial énfasis en los puntos clave durante su ejecución mediante un ejemplo hipotético.

© 2011 Sociedad Española de Cardiología. Publicado por Elsevier España, S.L. Todos los derechos reservados.

Palabras clave:

Revisión sistemática

Metaanálisis

Medicina basada en la evidencia

Abbreviations

MA: meta-analysis

RCT: randomized clinical trial

SR: systematic review

INTRODUCTION

It has been a tiring week. You sit down to look back calmly on the decisions you've made. These included requesting an operation for a patient with three-vessel disease, deciding whether to treat an 82 year old patient with inferior infarction of 70 min duration with fibrinolytic therapy or move her to your center for primary angioplasty, and deciding on anticoagulation treatment for an outpatient with atrial fibrillation.

* Corresponding author: Unidad de Epidemiología, Servicio de Cardiología, Hospital Vall d'Hebron, Pg. Vall d'Hebron 119-129, 08035 Barcelona, Spain.

E-mail address: nacho@ferreiragonzalez.com (I. Ferreira González).

Although you are reasonably sure your decisions were based on the best available evidence, you have some lingering doubts. Perhaps studies have been published which could lead your decisions to be questioned? Or perhaps different studies of the same intervention have produced different results? It's true that you have not had much time for reading over the past few months. To quickly clear up your doubts, you realize you need a concise, current, and rigorous summary of the best available evidence regarding the decisions you had to take. In other words, you need a systematic review (SR).¹

SRs are considered to be the most reliable source in informing medical decision-making,² which may explain their increasing popularity and the large rise in the number of SRs published in recent years.² However, performing a high-quality SR is not easy. There are rules governing the way they should be carried out and, as with other designs, recommendations on how results should be presented. These quality control guidelines have been developed by international, multidisciplinary groups of experts which include authors of SRs, methodologists, clinicians, and editors.²⁻⁴ This article presents the underlying rationale for performing and interpreting SRs and uses a hypothetical example to highlight key-points in their execution.

CONCEPT AND NOMENCLATURE

SRs are scientific investigations in which the unit of analysis is the original primary studies. These are used to answer a clearly formulated question of interest using a systematic and explicit process. For that reason, SRs are considered to be secondary research (“research-based research”). On the other hand, reviews which do not follow a systematic process (narrative reviews) cannot be considered to constitute a formal research process, but are simply a type of scientific literature based primarily on opinion.

From a formal point of view, SRs summarize the results of primary research using strategies to limit bias and random error.⁵ These strategies include:

- Systematic and exhaustive searching for all potentially relevant articles.
- The use of explicit and reproducible criteria to select articles which are eventually included in the review.¹
- Describing the design and implementation of the original studies, synthesizing the data, and interpreting the results.

Although SRs are a tool for synthesizing information, it is not always possible to present the results of the primary studies briefly. When results are not combined statistically, the SR is called a qualitative review. In contrast, a quantitative SR, or meta-analysis (MA) is an SR which uses statistical methods to combine the results of two or more studies.¹

An SR should not to be confused with an MA. The first is always possible, while the second is only sometimes possible. However, when conditions allow, MAs provide very useful, manageable information regarding the effect of a treatment or intervention, both in general and in specific patient groups. In addition, MAs make it possible to estimate the effect of an intervention more precisely and to detect moderate but clinically important effects that may have gone undetected in the primary studies. Typically, MAs combine aggregate data from published studies, but sometimes individual data from patients in different studies can be combined. This is called individual patient data meta-analysis and is considered the gold standard in SR.⁶

It should be noted that, in contrast to narrative reviews, SRs use a systematic method to search for all potentially relevant studies and apply explicit, reproducible, previously defined criteria to select the articles included in the final review. It is these features which give SRs their scientific character, in contrast to narrative reviews. Table 1 shows the difference between the two types of review.

As in clinical trials, a protocol should be developed prior to carrying out an SR.⁷ This will help the researchers to give due consideration to the most appropriate methods for use in the review and will also prevent decisions being taken *a posteriori* based on the results. The first international register of protocols for systematic reviews, apart from the Cochrane SRs, was recently

published under the name of PROSPERO (<http://www.crd.york.ac.uk/prospéro/>).

STAGES IN A SYSTEMATIC REVIEW

Briefly, a SR consists of the following steps:

- Definition of the clinical question of interest and the inclusion and exclusion criteria for studies.
- Identification and selection of relevant studies.
- Extraction of data from primary studies.
- Analysis and presentation of results.
- Interpretation of results.

Definition of the Clinical Question of Interest

The first step is to correctly formulate the clinical question of interest. In general, this should be explicit and structured so as to include the following key components:⁸

- The specific population and context. For example, elderly patients (over 75 years) admitted for acute myocardial infarction with ST elevation.
- The exposure of interest. This could be a risk factor, a prognostic factor, an intervention or treatment, or a diagnostic test. In the case of an intervention, treatment or diagnostic test a control exposure is usually defined at the same time. For example, primary angioplasty (intervention) versus fibrinolysis (control).
- Events of interest. For example, total mortality, cardiovascular mortality, readmission for acute coronary syndrome, revascularizations, etc.

From these elements, you might frame the question as follows: compared with fibrinolysis, does primary angioplasty reduce mortality and myocardial infarction in patients over 75 years of age? Once the question of interest has been defined and circumscribed, it is easier to establish the inclusion and exclusion criteria for primary studies. An ill-defined research question, on the other hand, leads to confused decision-making about which studies may be relevant in answering the question.

In many cases not easy to decide what the specific research question should be. It is however clear that it should be clinically relevant. If questions are too vague (e.g. is primary angioplasty useful in acute myocardial infarction?), they will be of little help to the clinician when making a decision about a particular patient. Exposures or patient characteristics which may affect the event of interest should also be taken into account. For example, it is not uncommon for patients over 75 years of age to be treated with oral anticoagulation, which could affect the expected event of interest. The study population could therefore be restricted to patients who are not receiving oral anticoagulation. However, overly specific inclusion criteria may limit the applicability of the results. Another

Table 1
Differences Between Systematic and Narrative Reviews.

Characteristic	Narrative review	Systematic review
Question of interest	Not structured, not specific	Structured question, well-defined clinical problem
Article search and sources	Not detailed and not systematic	Structured and explicit search
Selection of articles of interest	Not detailed and not reproducible	Selection based on explicit criteria uniformly applied to all articles
Assessing the quality of the information	Absent	Structured and explicit
Synthesis	Often a qualitative summary	Qualitative and quantitative summary
Inferences	Sometimes evidence-based	Normally evidence-based

option is to define a broad question that makes clinical sense and go on from there to explore more specific questions. For example, we could include all patients with acute myocardial infarction and then use exploratory analysis to determine the effect of the intervention in those treated with oral anticoagulants and those who are not. However, this strategy can lead to problems which are similar to those found in subgroup analysis.⁹ Finally, using inclusion criteria which are too broad means running the risk that the analysis will make no real clinical or biological sense.¹⁰

At this stage, it is important to decide which study designs to include in the review. The decision depends on the type of question we are trying to answer. If the idea is to evaluate the efficacy of an intervention, as in the previous example, we clearly need to include randomized clinical trials (RCT), if any are available. The same is true for the assessment of the reliability and safety of a diagnostic test. For SRs of community or public health interventions or when interventions are assessed over the long-term (particularly as regards their safety), then observational studies are more relevant. Occasionally, we may find that there are no RCTs available on a given intervention and, in that case as well, we will need to analyze observational studies.

Figure 1 shows, in simplified form, the elements of the question of interest in our example and the inclusion and exclusion criteria for the SR deriving from it.

Identifying and Selecting Studies

This stage consists of several parts:

1. Identifying potential articles
 - Deciding on restrictions regarding the publication language.
 - Deciding on the sources for obtaining primary studies.
 - Obtaining the titles and abstracts of potential primary studies.
2. Selecting potential articles
 - Applying the inclusion and exclusion criteria to the titles and abstracts obtained.
 - Obtaining potential articles based on eligible titles and abstracts and applying the inclusion and exclusion criteria.
 - Evaluating the level of agreement in study selection.

Identifying Potential Articles

As with any research study, the presence of errors in data extraction can invalidate the results of a SR. So as to minimize random error and bias, it is extremely important to obtain as many

primary studies as possible on the question of interest. If studies are omitted, it can introduce bias if the final sample of selected articles is not representative. Two decisions need to be made at this point: firstly, whether to restrict the language of publication and whether to include studies which are not published in medical journals.

With regard to language, for practical purposes only publications in English and the native language of the author of the SR are usually included. Nevertheless, studies have shown that research quality is not necessarily related to the language of publication.¹¹ Restricting the language of publication may also bias the results of the SR by excluding potentially relevant studies.¹²

For reasons of practicality, it seems reasonable to include only studies published in medical journals. One could argue that the fact that these studies have been through a peer review process makes them the most reliable.¹³ However, independently of quality, it is also true that studies with negative or inconclusive results are less likely to be published,^{14,15} and their exclusion may bias the results of the SR. This is called publication bias and it makes it more likely that a SR which excludes unpublished studies will overestimate the relationship between exposure and the event of interest. In extreme situations, this bias can lead to completely futile treatments being presented as effective.^{16,17}

Once a decision has been taken on the preceding two questions, the following step is crucial: where to search for primary studies? Various strategies are available:

- Electronic databases: MEDLINE, EMBASE, CENTRAL.
- Non-indexed databases: AMED, CINAHL, BIOSIS, etc.
- Hand-searching of journal content tables, minutes and summaries of scientific meetings, and books.
- Lists of references and citations: Science Citation Index and similar.
- Records of ongoing studies (e.g., clinicaltrials.gov).
- Contact with pharmaceutical companies.
- Contact with fellow experts in the field of interest.

For obvious reasons, the most widely used strategy today is to perform the search in electronic databases. This is not, however, a simple strategy. Although there is some overlap between databases, many journals included in one database are not included in others. MEDLINE, for example, only indexes about 5600 of the more than 16 000 existing biomedical journals, and most of those are in English. EMBASE indexes over 1000 journals which are not included in MEDLINE, many of them European. Other databases further complement MEDLINE and EMBASE, as efforts have been made to register literature on unpublished studies (gray literature).¹⁸

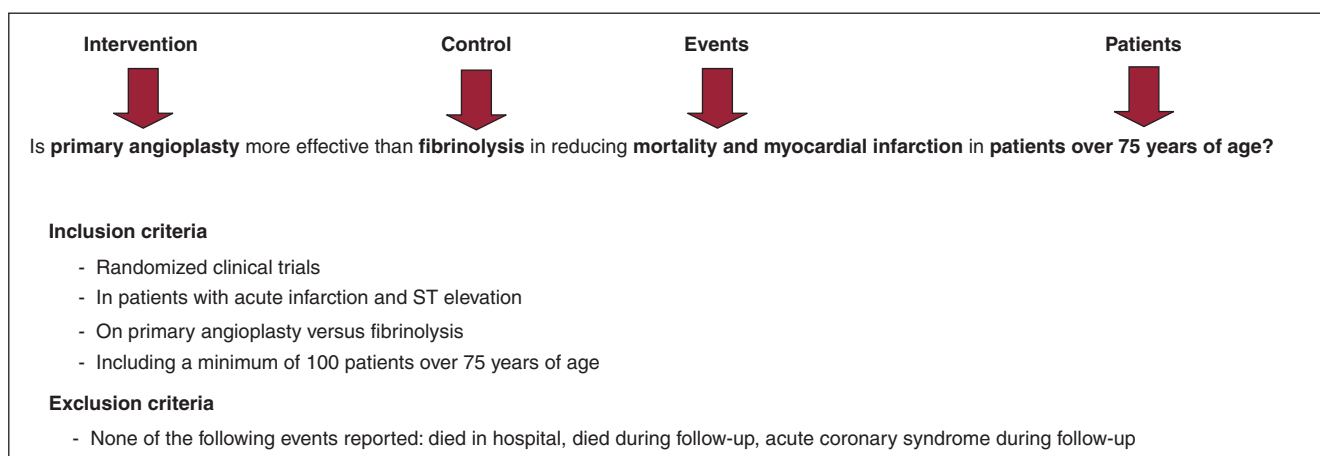


Figure 1. Example of the elements making up the question of interest for a systematic review, and inclusion and exclusion criteria for the primary studies.

For practical reasons, the identification of potentially eligible articles is often undertaken by identifying titles and abstracts. However, each database has a particular structure and uses more or less specific criteria for indexing and keywords. For example, in MEDLINE, the specific index of controlled vocabulary is called MeSH (Medical Subject Headings). The assistance of a documentalist who is expert in SRs is crucial at this stage. The final search strategy should also be described in the publication of the SR, to verify its reproducibility.

Depending on the topic of interest, restricting the search for studies exclusively to electronic databases may not be optimal. Sometimes it is convenient to include a complementary strategy to identify unpublished studies. This phase is one of the most laborious and often requires hand searching of journal abstracts or conference proceedings, contact with experts in the field or with pharmaceutical companies, etc. Clearly, this involves a significant investment of time and financial resources. Recognizing this, the Cochrane Collaboration has led an international initiative to develop a register of controlled trials, formerly known as the Cochrane Controlled Trials Register, and now called CENTRAL.¹⁹ This resource, which is constantly updated, contains hundreds of thousands of records or citations of studies published in indexed journals as well as supplements (usually conference proceedings) and is of undoubted interest in identifying RCTs.²⁰

Applying search strategies in several electronic databases usually generates a large number of references many of which are likely to be duplicated between databases. It is therefore useful at this stage to use software for the automated management of bibliographic citations, such as ProCite or Reference Manager.

Selection of Potential Articles

Once we have a first list of titles and abstracts, we need to make an initial selection or screening of potentially eligible articles. A study selection form based on explicit, understandable criteria is useful at this stage.²¹ A simple example for the research question referred to above is shown in Figure 2. The selection process usually begins with a review of titles and abstracts; if there are doubts about the suitability of an article, the full text should be reviewed.

In order to increase the reliability and thoroughness of the process, study selection should be carried out by two independent reviewers. It is equally important to measure the degree of agreement between reviewers by calculating the kappa statistic for each of the items on the selection form. Simply put, this statistic measures the degree of agreement between reviewers above that expected by chance.^{22,23} Where reviewers disagree about whether to include an article or not, a third, senior researcher is usually appointed to act as arbitrator and to make the final decision.

Finally, it is important to note that the entire process of identifying and selecting studies should be accurately reported using a flow diagram of articles identified in each phase, articles eliminated, and the reasons for their elimination (Fig. 3).

Extracting Data From Primary Studies

In this phase, information should be extracted from each study as reliably as possible and should therefore be done in duplicate if

<p>Study reference: _____</p> <p>Selection level (check all that apply): Title ___ Abstract___ Text___</p> <p>Selection criteria:</p> <p>Patients Are patients with acute infarction and elevated ST studied? Yes/No Are patients over 75 years of age included? Yes/No</p> <p>Intervention Did at least one of the treatment groups receive primary angioplasty? Yes/No</p> <p>Control Did at least one of the treatment groups receive fibrinolysis? Yes/No</p> <p>Events Are any of the following events included as individual or combined outcome variables: overall mortality, cardiovascular mortality, non-fatal myocardial infarction, new revascularizations? Yes/No</p> <p>Design Is it a randomized clinical trial? Yes/No</p> <p>ACTION (ONLY include if the answer to ALL of the above was 'yes'): Include___ Exclude___ Doubtful___</p>
--

Figure 2. Hypothetical example of a form for selecting potential articles for a systematic review.

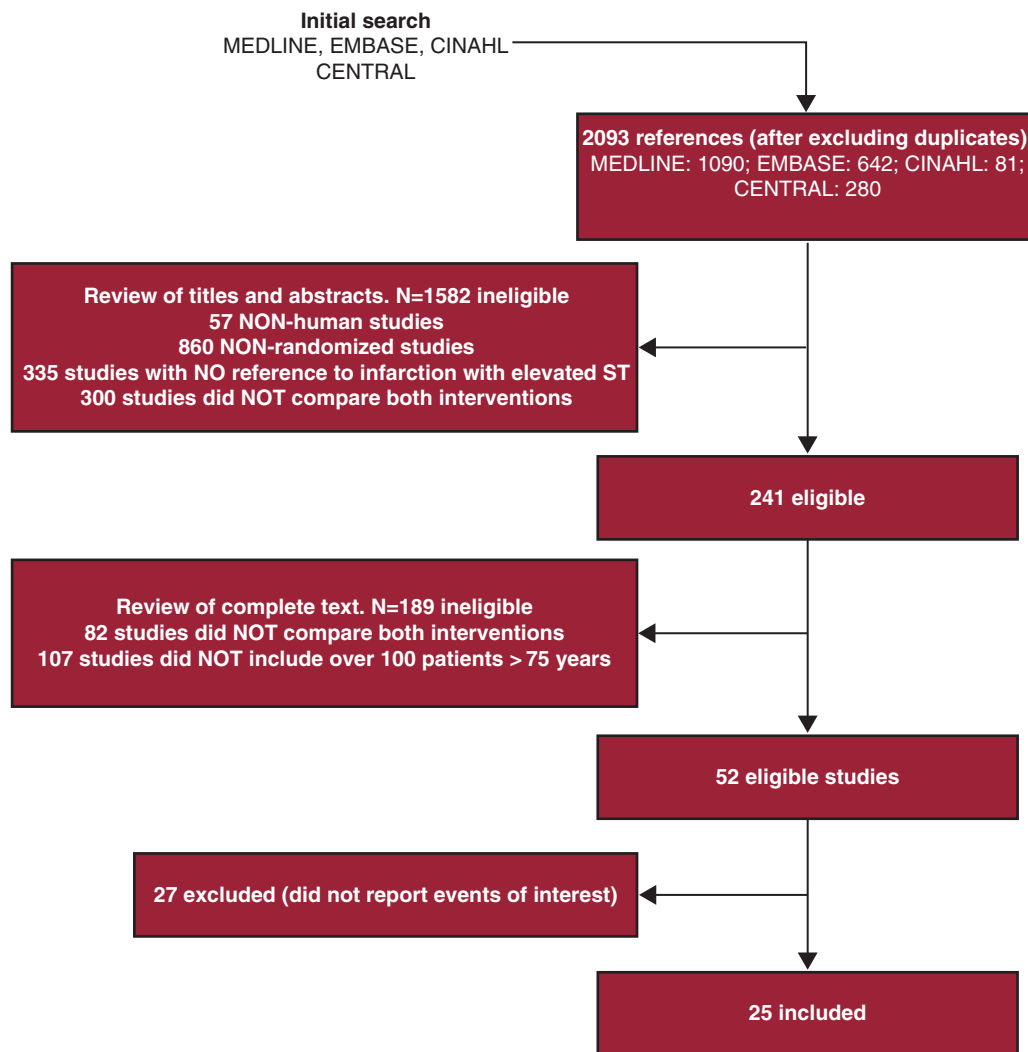


Figure 3. Example of the article selection process.

possible. If that is not feasible, an alternative is to have another reviewer carry out an independent audit of a random sample of studies.

The information to be extracted from the primary studies should be agreed on during the design phase of the study. In general, it should be data that allows the study hypothesis to be accepted or rejected. In summary, the data extraction form usually includes:²⁴

- A. Information on patients, the intervention of interest, the control intervention, and study design.
- B. Information on the results.
- C. Information on the methodological quality of the study.

Point A covers all information that may impact the outcome, and may vary between studies. In our example, an RCT which included patients with a mean age of 65 years and a median door-to-balloon time of 63 min would not be the same as another trial in which patient mean age was 75 years and door-to-balloon time was 96 min. All of these differences could explain a difference in the magnitude of the intervention effect between studies, i.e. they could help to explain the heterogeneity of the effect. The challenge is to strike a balance between the completeness of the information collected while, at the same time, avoiding

including too much unnecessary information that would overload the review.

The information in B corresponds to the extraction of results. The format chosen will depend on how the events of interest are defined: whether as a dichotomous variable or as a continuous variable. In the first case, it will sometimes be easy to get the number and percentage of patients in both branches of the study in which the event of interest occurred. On other occasions, the results may be expressed as a measure of association or impact, such as relative risk, relative risk reduction, absolute risk reduction, odds ratio or hazard ratio. Figure 4 shows an example of a hypothetical form for the extraction of results. Fortunately, it is easy to get the number and approximate percentage of patients with the event of interest from any of the usual measures of association or impact. In the event that the outcome variable is continuous, such as the ejection fraction, the information of interest is the mean and standard deviation in each treatment group.

Finally, the data extraction form should include information on the methodological quality of the studies included, as this is closely related to the magnitude of the effect. There is controversy regarding the best way to reflect the methodological quality of a study.²⁵ Some argue for the use of quality rating scales and several of these have been developed,²⁶ mostly for RCTs. Some are generic

Study reference: _____

	Primary angioplasty Total no randomized: _____ Total no analyzed: _____		Fibrinolysis Total no randomized: _____ Total no analyzed: _____		Measures of association or impact			
	N	%	N	%	RR (95% CI)	RRR (95% CI)	ARR (95% CI)	HR (95% CI)
Total mortality								
CV mortality								
Readmissions for ACS								
New revascularizations								

Figure 4. Hypothetical example of a form for extracting results from studies included in a systematic review. ACS, acute coronary syndrome; ARR, absolute risk reduction; CI, confidence interval; CV, cardiovascular; HR, hazard ratio; RR, relative risk; RRR, relative risk reduction.

and others specific to certain clinical areas. However, it has been shown that using one scale or another can lead to substantial variation in the results of an MA,²⁷ so that none of them are totally reliable. Recently, a new system to measure the methodological quality of studies included in SRs has been developed called GRADE (Grading of Recommendations Assessment, Development, and Evaluation).²⁸ This system, which was developed and agreed upon by a group of international leaders in the implementation of clinical practice guidelines, offers some interesting advantages over the others. Basically:

- The quality of the evidence, which is classified as high, moderate, low and very low, is reported separately from the grade of recommendation (strong or weak recommendation).
- The values and preferences of patients are recognized and incorporated.
- It provides a clear and pragmatic level of recommendation (strong or weak) for clinicians, patients and managers.
- It explicitly assesses the importance to patients of the outcome variables for the therapeutic alternatives considered

Some authors advocate collecting and evaluating the methodological elements of each individual study, instead of using the controversial scales.²⁹ The information collected will depend on the study design. In the case of RCTs, which are the most common design in reviews of interventions, aspects of design and implementation most frequently related to the risk of bias are: the way randomization is concealed, the method of masking interventions, tracking losses, the type of analysis, or early termination of the trial because of an apparent benefit. **Figure 5** shows a hypothetical example of a data collection form for methodological elements of RCTs.

Analysis and Presentation of Results

Summarizing and presenting the results of primary studies obtained using a systematic, reproducible methodology constitutes, in itself, a qualitative SR. The next step is to combine the results of primary studies using statistical methods, i.e., meta-analysis (MA).

Conceptually, MAs are used to combine the results of two or more similar studies on a particular intervention, as long as the same outcome variables are used. Meta-analysis does not provide a simple arithmetic average of the results of different studies, but a weighted average. In other words, it gives greater weight to studies with a larger information load, i.e. studies that are larger and / or

have a higher number of events. On combining the results, a different weight is assigned to each study and it is that which provides the weighted average. Because MA also takes into account intra- and inter-study variability when combining results, the validity of the conclusions is further enhanced. If there is a large amount of variation between the results of the studies included (heterogeneity) it may not be appropriate to combine the results statistically. In that case, only the results and the characteristics of individual studies should be presented, without further statistical treatment. It can be useful to present the results graphically and the study characteristics in tables.

Basically, two types of models can be used to statistically combine results, those being fixed effects models and random effects models. The first assumes that the treatment effect is constant in all studies, while the second assumes that the effect is randomly distributed among the different studies. In other words, the fixed effects model assumes that there is only one source of variability in the results (the study), while the random effects model introduces a second source of variation between studies. The practical consequence of this is that the random effects model tends to produce more conservative estimates (wider confidence intervals) of the combined effect. Which model we use depends on our view of similarities and differences between the studies to be merged, although it is common to use both approaches.

But what do we mean when we say that there is “heterogeneity” between studies, and how is that heterogeneity measured? Basically, it means that, after weighting the studies, the results of individual studies (the effect of the intervention) differ more than would be expected by chance. In other words, the effect of the intervention “differed” between studies either because of differences in the study design used, in the methods used to collect information, or the type of analysis used, and/or the characteristics of the study population. For example, suppose that one clinical trial of an antihypertensive drug included 70% African Americans and another study, using the same drug, included only 10%. Now imagine that the drug has a very powerful antihypertensive effect in African American populations and no effect in the rest of the population. Clearly, the first study will show the intervention has a positive effect while the second study will show it has no effect. It would not be appropriate to combine the results of the two studies because it would show a “mean” effect for the intervention which would hide a much richer and complex reality, and would lead to confusion.

Various statistical tests are available to quantify heterogeneity. The most common are the Q, H and I² statistics. The easiest to interpret is the I², which indicates the proportion of observed

<p>1. Allocation concealment of random assignment:</p> <ul style="list-style-type: none"> • Central randomization (via telephone call, Web page, independent investigator or through the pharmacy) • Coded containers with the study and control medication • Sequentially-numbered, opaque, sealed envelopes • Envelopes used, but without specifying the three above-mentioned elements • Pseudo-randomization (coin toss, birth-date, day of the week, etc.) • Masking of randomization mentioned but not described • Not reported • Lack of randomization masking is reported <p>2. Masking or blinding of the intervention (check that which applies):</p> <p>Patients</p> <ul style="list-style-type: none"> • Health care personnel administering the medication • Investigators collecting the data • Investigators evaluating events • Investigators carrying out the statistical analysis <p>3. Participants lost to follow-up:</p> <p>N= _____ %= _____</p> <ul style="list-style-type: none"> • Not reported <p>4. Data analysis</p> <p>Intention to treat (patients are analyzed in the intervention arm to which they were randomized)</p> <ul style="list-style-type: none"> • Per protocol (patients are analyzed according to the intervention they eventually received) • Both • Not reported <p>5. The study was interrupted early because of a benefit</p> <ul style="list-style-type: none"> • Yes • No

Figure 5. Hypothetical example of a form to assess the quality of randomized clinical trials included in a systematic review.

(between-study) variability in the intervention effect that is due to heterogeneity between studies, i.e. non-random variability. It is generally held that a proportion of 25% indicates that there is little heterogeneity, 50% shows moderate heterogeneity, and 75% a high degree of heterogeneity.³⁰

The results of MAs are presented graphically using forest plots. This type of graph shows data from individual studies together with a representation of the statistical weight of each study in relation to confidence intervals and standard error of the mean. For example, suppose that in our example we combine eight studies to analyze the effect on cardiovascular mortality of angioplasty compared to fibrinolysis in patients aged over 75 years. Figure 6A presents the combined analysis of eight studies in the form of a forest plot. It can be seen that the overall effect is inconclusive and, moreover, that there is significant heterogeneity between studies ($I^2 = 90\%$).

A simplistic interpretation would be that primary angioplasty is not superior to fibrinolysis in reducing cardiovascular death in patients over 75 years of age. However, the high heterogeneity between studies should make us suspect that this result is unrealistic. In fact, visually it can be seen that angioplasty showed a superior effect to fibrinolysis in studies 2, 4, 5 and 6, but was

inferior in studies 1, 3, 7 and 8. A careful reading of these surveys indicates that mean door-to-balloon time was <90 min in all studies in which angioplasty showed a superior effect, but that it was inferior to fibrinolysis in all studies with a door-to-balloon time over 140 mins. In this case it would be justifiable to perform a subgroup analysis in which studies with a door-to-balloon time <90 min and those with longer door-to-balloon times are analyzed separately. Figure 6B shows the results. It can be seen that heterogeneity is significantly reduced in each subgroup and that the results are consistent, with primary angioplasty showing a significant benefit for studies with lower door-to-balloon times (bottom of Fig. 6B) but a smaller effect than fibrinolysis in studies with longer door-to-balloon times (top of fig. 6B).

Finally, it is important to point out that guidelines and recommendations for the presentation and publication of SR are available and provide checklists of items to be covered in the publication. The MOOSE guidelines detail the specific items that SRs of observational studies should include,⁴ while the PRISMA guidelines (which replaced the QUORUM recommendations) cover MAs of RCTs.³ They can be consulted on-line at <http://www.consort-statement.org>

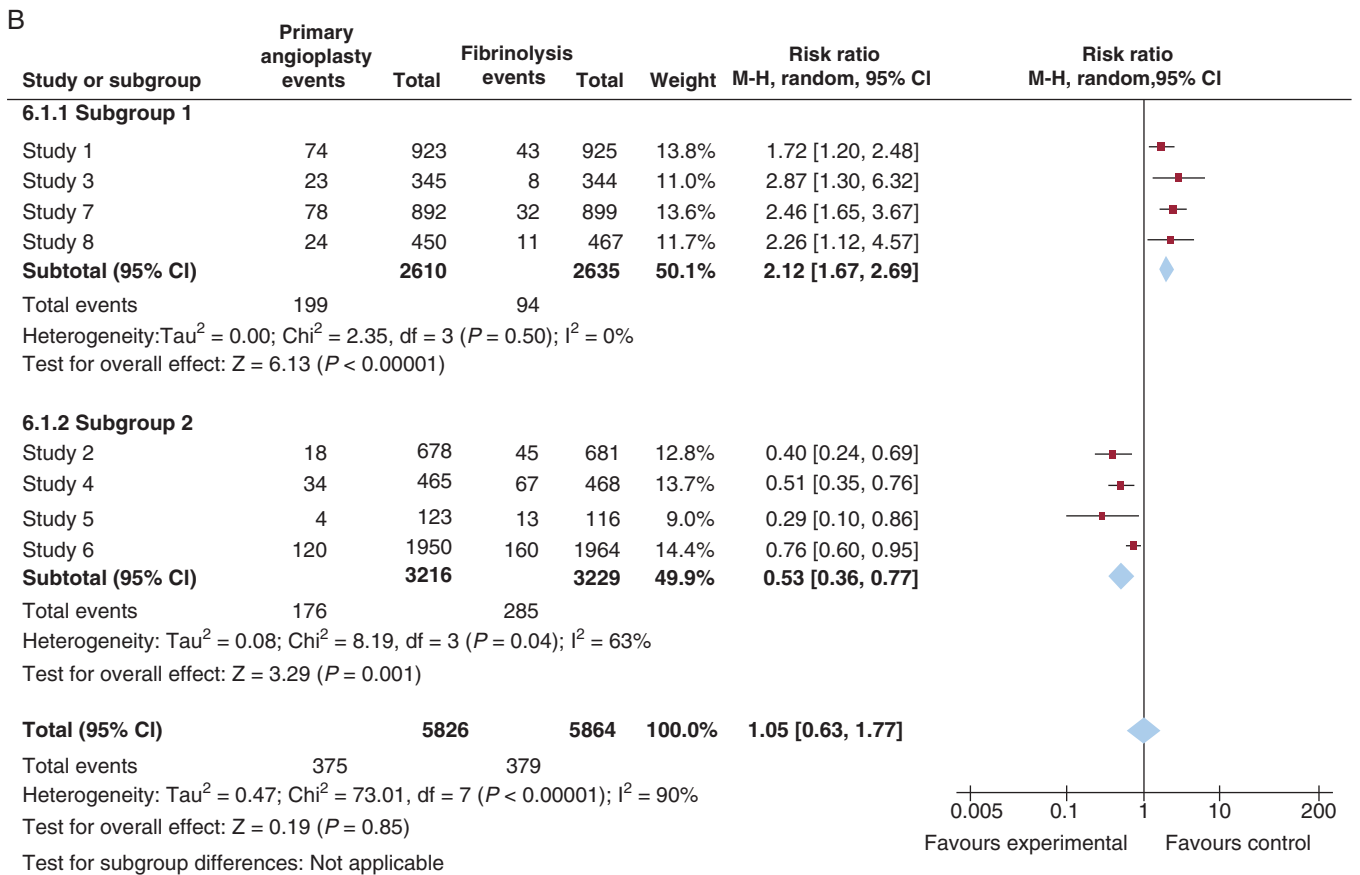
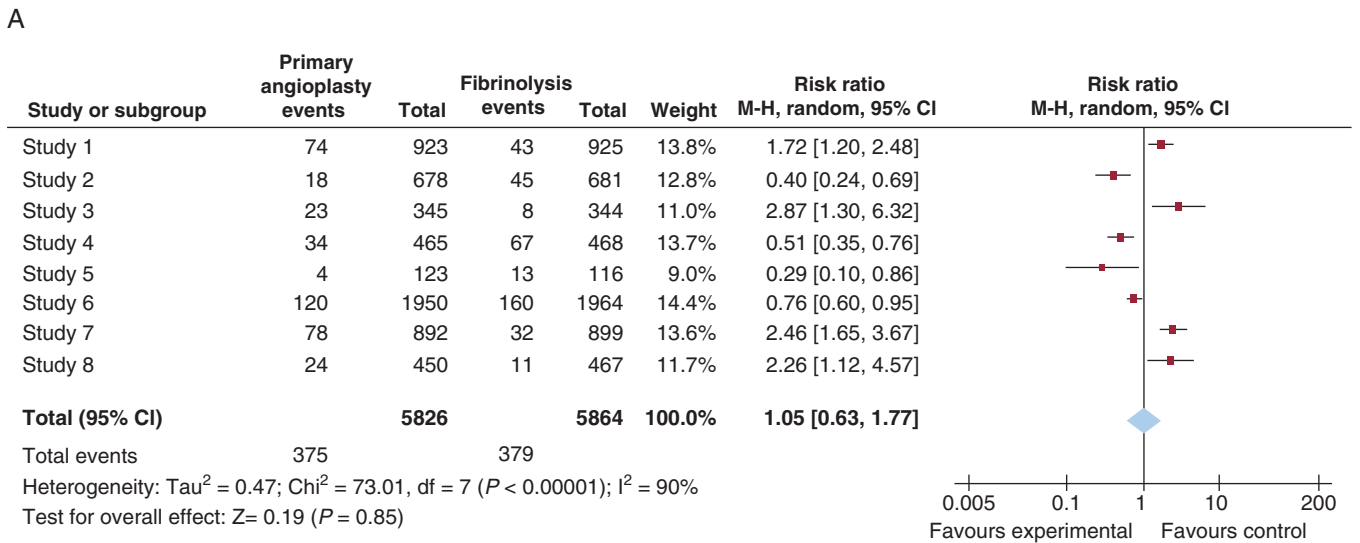


Figure 6. Meta-analysis of the effect of primary angioplasty versus fibrinolysis on total mortality in patients with myocardial infarction. A, including all studies. B, analysis of subgroups based on door-to-balloon time; upper figure, studies in which mean door-to-balloon time exceeded 140 min; lower figure, studies in which the mean door-to-balloon time was <90 min. CI: confidence interval.

Interpreting the Results

Finally, the SR ends with the interpretation of results. This includes a discussion of the limitations of the study (of the review) as well as any potential biases of the original studies, and potential biases that could affect the SR itself. It is also important to discuss the consistency of the findings and their applicability, and to propose recommendations for future research on the topic of interest.

SYSTEMATIC REVIEWS AND CLINICAL PRACTICE

As much as the findings of an RS are consistent and convincing, in the end it is the clinician who must make decisions regarding a particular patient. The findings of a SR should not be taken as fixed and unchanging standards resulting from an “evidentialist orthodoxy”. In other words, we must adapt the findings of an SR to the patient, and not vice versa. In this regard, before making a

decision on a patient based on an SR, we recommend making the following reflections.³¹

Are the Findings Applicable to My Patient?

The SR has shown that primary angioplasty is superior to fibrinolysis in elderly patients. But imagine my particular patient has a condition that was an exclusion criterion in all clinical trials included in the SR, such as a creatinine clearance rate <30 ml / min. In this case, my patient would not be represented by the specific clinical trials included in the SR.

Is the Intervention Feasible in My Patient?

There may be regional differences in the availability and/or experience of applying a particular technique, differences which should be taken into account when applying the intervention to an individual patient.

What is the Risk-Benefit for My Patient?

Even if the intervention is feasible and applicable, the specific risks for a particular patient should be taken into account and this is an aspect which is generally poorly represented by clinical trials.

What Are My Patient's Particular Values and Preferences?

Accustomed to making decisions based markers of myocardial necrosis, electrocardiograms, and other complementary tests, we can succumb to an overly paternalistic form of medicine, whereby we treat the patient with the best intentions, but without taking his or her point of view into account.

CONCLUSIONS

SRs are an essential tool in synthesizing available scientific information; they enhance the validity of the findings from individual studies and identify areas of uncertainty where research is needed. They are essential to the practice of evidence-based medicine. However, SRs should be implemented following a strict methodology and applying quality control to avoid biased conclusions. Ultimately, it is the clinician who must make decisions regarding a particular patient; SRs are just one more tool to be used, judiciously, in making those decisions.

CONFLICTS OF INTEREST

None declared.

REFERENCES

- Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for clinical decisions. *Ann Intern Med.* 1997;126:376–80.
- Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and reporting characteristics of systematic reviews. *PLoS Med.* 2007;4:e78.
- Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *J Clin Epidemiol.* 2009;62:e1–34.
- Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA.* 2000;283:2008–12.
- Cook DJ, Sackett DL, Spitzer WO. Methodologic guidelines for systematic reviews of randomized control trials in health care from the Potsdam Consultation on Meta-Analysis. *J Clin Epidemiol.* 1995;48:167–71.
- Stewart LA, Tierney JF, Clarke M. Reviews of individual patient data. In: Higgins JP, Green S, editors. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.1* (updated September 2008). The Cochrane Collaboration; 2008.
- Straus S, Moher D. Registering systematic reviews. *CMAJ.* 2010;182:13–4.
- Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions. *ACP J Club.* 1995;123:A12–3.
- Sun X, Briel M, Walter SD, Guyatt GH. Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. *BMJ.* 2010;340:c117.
- Dickersin K, Berlin JA. Meta-analysis: state-of-the-science. *Epidemiol Rev.* 1992;14:154–76.
- Moher D, Fortin P, Jadad AR, Juni P, Klassen T, Le Lorier J, et al. Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews. *Lancet.* 1996;347:363–6.
- Gregoire G, Derderian F, Le Lorier J. Selecting the language of the publications included in a meta-analysis: is there a Tower of Babel bias? *J Clin Epidemiol.* 1995;48:159–63.
- Chalmers TC, Levin H, Sacks HS, Reitman D, Berrier J, Nagalingam R. Meta-analysis of clinical trials as a scientific discipline. I: Control of bias and comparison with large co-operative trials. *Stat Med.* 1987;6:315–28.
- Dickersin K, Min YI. Publication bias: the problem that won't go away. *Ann N Y Acad Sci.* 1993;703:135–46.
- Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet.* 1991;337:867–72.
- Simes RJ. Confronting publication bias: a cohort design for meta-analysis. *Stat Med.* 1987;6:11–29.
- Egger M, Smith GD. Misleading meta-analysis. *BMJ.* 1995;311:753–4.
- Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. *BMJ.* 1994;309:1286–91.
- Dickersin K, Larsen K. Section V: Establishing and maintaining an International register of RCTs. In: *The Cochrane Collaboration Handbook*. London: BMJ Publishing Group; 2010.
- Royle P, Milne R. Literature searching for randomized controlled trials used in Cochrane reviews: rapid versus exhaustive searches. *Int J Technol Assess Health Care.* 2003;19:591–603.
- Meade MO, Richardson WS. Selecting and appraising studies for a systematic review. *Ann Intern Med.* 1997;127:531–7.
- Streiner DL, Norman GR. Reliability. In: *Health measurement scales: a practical guide to their development and use*. New York: Oxford University Press; 1995.
- Altman DG. Measuring agreement. In: *Practical statistics for medical research*. London: Chapman and Hall; 1991.
- Guyatt GH, Rennie D, Meade MO. Summarizing the evidence. In: *Users' guides to the medical literature*. New York: McGraw Hill; 2009.
- Moher D, Jadad AR, Tugwell P. Assessing the quality of randomized controlled trials. Current issues and future directions. *Int J Technol Assess Health Care.* 1996;12:195–208.
- Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Control Clin Trials.* 1995;16:62–73.
- Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA.* 1999;282:1054–60.
- Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ.* 2008;336:924–6.
- Juni P, Altman DG, Egger M. Assessing the quality of randomized controlled trials. In: Egger M, Smith GD, Altman DG, editors. *Systematic reviews in health care: metaanalysis in context*. London: BMJ Publishing Group; 2001.
- Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med.* 2002;21:1539–58.
- McAlister FA. Applying the results of systematic reviews at the bedside. In: Egger M, Smith GD, Altman DG, editors. *Systematic reviews in health care: meta-analysis in context*. London: BMJ Publishing Group; 2001.

PUBLICACIÓN CIENTÍFICA BIOMÉDICA

Cómo escribir y publicar
un artículo de investigación

J. Jiménez Villa
J.M. Argimon Pallàs
A. Martín Zurro
M. Vilardell Tarrés



Ámsterdam Barcelona Beijing Boston Filadelfia Londres Madrid
México Milán Múnich Orlando París Roma Sidney Tokio Toronto

Índice de capítulos

Autores	IX
Presentación	XV
1. La literatura científica biomédica	1
<i>J.M. Argimon, J. Jiménez, A. Martín Zurro y M. Vilardell</i>	
PARTE I. ELABORACIÓN DE UN MANUSCRITO	
2. Preparativos para la redacción de un manuscrito	17
<i>F. García Río</i>	
3. Estructura: el artículo original y otros tipos de artículo	37
<i>J.M. Argimon</i>	
4. Presentación de resultados estadísticos y elaboración de tablas	57
<i>J. Jiménez</i>	
5. El impacto de los gráficos en la comunicación científica	79
<i>J.A. González y L. Jover</i>	
6. El estilo científico	97
<i>M. Pulido</i>	
7. Escribir en inglés: normas generales e interacción con el traductor ...	109
<i>X. Bosch</i>	
PARTE II. CONTENIDO DE UN ARTÍCULO ORIGINAL	
8. Un paseo estadístico por las guías de publicación	125
<i>J. Cortés y E. Cobo</i>	

9.	Ensayos clínicos aleatorizados	145
	<i>A. Cobos</i>	
10.	Ensayos de equivalencia y no inferioridad	161
	<i>J. Jiménez y J.M. Argimon</i>	
11.	Ensayos clínicos con asignación aleatoria de grupos	175
	<i>J.M. Argimon</i>	
12.	Otras extensiones de la declaración CONSORT: intervenciones no farmacológicas, ensayos pragmáticos y resúmenes de ensayos clínicos	185
	<i>J. Jiménez</i>	
13.	Estudios epidemiológicos	205
	<i>E. Fernández</i>	
14.	Estudios de precisión diagnóstica y de pronóstico: las directrices STARD y REMARK	215
	<i>D. Rigau, P. Alonso e I. Solà</i>	
15.	Revisiones sistemáticas	231
	<i>G. Urrútia y X. Bonfill</i>	
16.	Calidad de vida relacionada con la salud y otros resultados percibidos por los pacientes	249
	<i>J. M. Valderas y J. Alonso</i>	
17.	La red EQUATOR: facilitando la comunicación transparente y precisa de la investigación sanitaria	265
	<i>I. Simera, D.G. Altman, D. Moher, K.F. Schulz y J. Hoey</i>	
PARTE III. DEL ENVÍO DEL MANUSCRITO A SU PUBLICACIÓN Y DIFUSIÓN		
18.	Requisitos para el envío de un manuscrito	273
	<i>A. Marušić</i>	
19.	El sistema de revisión por expertos (<i>peer review</i>) en las revistas científicas: ventajas y limitaciones	291
	<i>S. Lorenzo y G. Carrasco</i>	
20.	El proceso editorial	309
	<i>F. Cardellach y J.M. Ribera</i>	
21.	Autoría	327
	<i>J. P. Gisbert y J. M. Piqué</i>	

22.	Mala conducta y conflicto de intereses en las publicaciones científicas	347
	<i>A. Pascual y L. Martínez Martínez</i>	
23.	Bibliometría e indicadores de actividad científica	363
	<i>R. Alexandre</i>	
24.	Las revistas electrónicas	385
	<i>C. González y M.F. Abad</i>	
	Índice alfabético	399

Revisiones sistemáticas

G. Urrútia y X. Bonfill

IMPORTANCIA DE LAS REVISIONES SISTEMÁTICAS

El gran número de publicaciones científicas existentes actualmente pone de manifiesto la necesidad de gestionar de manera eficiente el conocimiento científico. Para hacernos una idea, solamente en la base bibliográfica Medline se hallan indexadas más de cinco mil revistas con más de dieciocho millones de artículos publicados. No solamente nos enfrentamos a una sobresaturación de información científica, sino que la calidad de la misma es muy variable, lo que nos obliga a realizar una valoración crítica cuidadosa antes de aplicarla. No obstante, el tiempo disponible para la obtención y valoración de los estudios suele ser muy limitado. Las revisiones de la literatura científica pueden ayudar a proporcionar información científica de forma rápida, en la medida que resumen en un único documento la información procedente de múltiples estudios individuales. Por este motivo las revisiones han experimentado un auge considerable en los últimos años y se han constituido como una pieza fundamental para la toma de decisiones informadas¹. Si en el año 1995 el número de artículos indexados en PubMed como «revisión» era de 48.747, este número aumentó a 64.986 en 2000, y a 85.883 en 2007. Se estima que actualmente se publican unas 2.500 revisiones cada año solamente en Medline, lo que las convierte en una parte esencial de la literatura biomédica².

Sin embargo, no todas las revisiones cumplen adecuadamente esta función. Así, las revisiones narrativas, habitualmente basadas en la opinión, suelen estar escritas por expertos en una determinada área científica, por lo que pueden estar sesgadas por dos razones principales: la posible dependencia de la subjetividad del revisor y la falta de explicitación de los métodos y los criterios de elaboración de la revisión³. Esto no significa que las conclusiones de una revisión narrativa nunca sean correctas, sino que las limitaciones inherentes a su formato impiden al lector conocer cuál es su grado de fiabilidad. De ahí la necesidad de disponer de instrumentos más fiables y precisos, como las revisiones sistemáticas, que nos permitan la toma de decisiones sanitarias sobre una base más sólida de información. Las revisiones sistemáticas se definen como revisiones de la evidencia existente para responder una pregunta bien delimitada, que utilizan métodos específicos y sistemáticos para identi-

ficar, seleccionar y evaluar críticamente la investigación primaria, así como para extraer y analizar los datos de los estudios que se incluyen en la revisión⁴.

La revisión sistemática se puede aplicar a cualquier pregunta de investigación, ya sea sobre la etiología, el diagnóstico o el pronóstico, o de intervención (evaluación del efecto de un tratamiento o intervención), aunque es en estas últimas donde más se han utilizado las revisiones sistemáticas con metaanálisis.

Las revisiones sistemáticas pueden incluir o no técnicas de metaanálisis. Los metaanálisis se definen como la utilización de técnicas estadísticas para combinar los datos de los estudios que han intentado responder la misma pregunta y presentarlos y resumirlos de manera gráfica y sintética (fig. 15-1). Aunque a menudo los términos «revisión sistemática» y «metaanálisis» se utilicen indistintamente, no son equivalentes: la revisión sistemática consiste en el proceso llevado a cabo para obtener conclusiones a partir de estudios originales, ya sea combinando estadísticamente o no sus resultados, mientras que el metaanálisis es la combinación matemática de al menos dos estudios para obtener una estimación global del efecto de la intervención que se está evaluando. Mientras que la revisión sistemática siempre es posible, el metaanálisis sólo lo es cuando se cumplen determinadas circunstancias. Cuando la revisión sistemática incluye un metaanálisis, la información proporcionada, además de otras ventajas, presenta la cualidad de ser sintética cuantitativamente, proporcionando un valor numérico que resume el efecto aproximado de la intervención analizada. El lector no sólo desea conocer si un tratamiento o intervención es efectivo o no y en qué condiciones, sino que también necesita saber la magnitud del efecto que cabría esperar cuando se aplica.

El valor creciente de las revisiones sistemáticas se pone de manifiesto en el trabajo de la Colaboración Cochrane, el primer propósito de la cual es generar, difundir y actualizar revisiones sistemáticas de alta calidad sobre los efectos de las intervenciones sanitarias. Esta organización internacional sin ánimo de lucro, que comenzó su andadura en el año 1993, publicó por primera vez en el año 1995 su *Cochrane Handbook for Systematic Reviews of Interventions* (Manual del Revisor Cochrane)⁵, cuya finalidad principal es el ayudar a los autores de revisiones Cochrane a ser sistemáticos y explícitos en el desarrollo de las mismas.

Si bien la mayoría de revisiones sistemáticas de la Colaboración Cochrane se basan en ensayos clínicos aleatorizados (ECA), en algunas ocasiones incluyen también otro tipo de diseños (estudios cuasi experimentales y estudios controlados no aleatorizados), especialmente cuando se trata de intervenciones complejas o del ámbito de la salud pública. Un buen ejemplo de ello es el grupo de Práctica Efectiva y Organización de la Atención Sanitaria (*Effective Practice and Organisation of Care Group*), focalizado en intervenciones educativas, conductuales, financieras, regulatorias u organizacionales diseñadas para mejorar la práctica profesional y el funcionamiento de los servicios en salud. La información sobre el grupo y sus métodos puede encontrarse en <http://www.epoc.cochrane.org/en/index.html>.

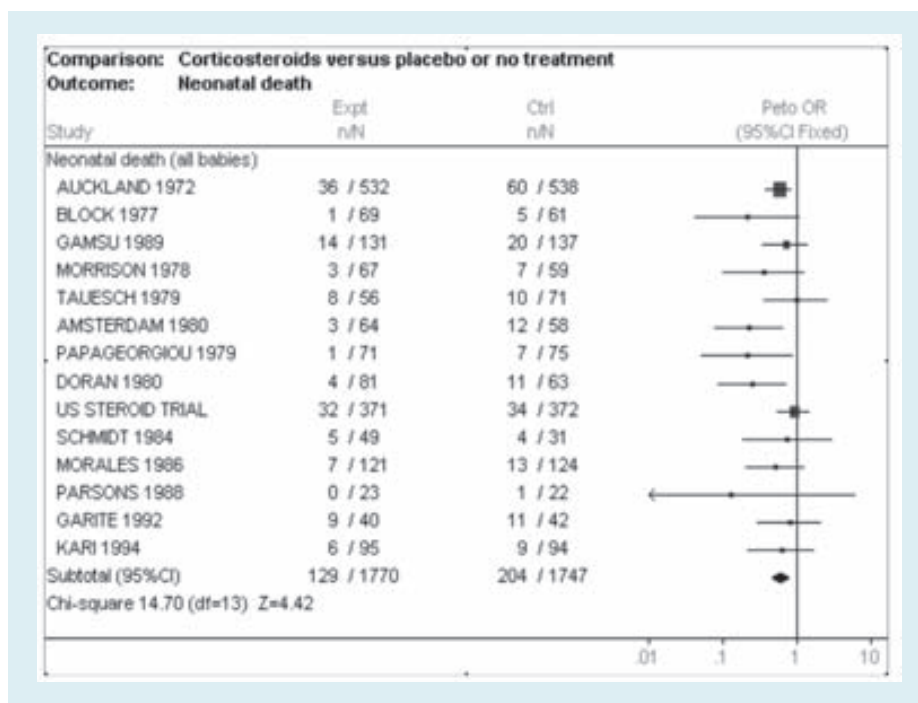


Figura 15-1. Esquema de un metaanálisis: «Corticoides para el parto prematuro»

Explicación del gráfico:

La primera versión estructurada publicada en 1989 de esta revisión constituye el logro de la Colaboración Cochrane. El objetivo de esta revisión fue evaluar la efectividad y seguridad a corto y largo plazo de la administración de corticoides en mujeres gestantes con la finalidad de acelerar la maduración pulmonar fetal antes de un parto prematuro.

En la columna de la izquierda se encuentra la lista de los estudios (ECA) de los que se han tomado los datos para el metaanálisis. La segunda columna, con el encabezamiento «Expt», contiene los datos del grupo experimental en cada estudio (mujeres gestantes que recibieron corticoides). Los valores indican el número de neonatos fallecidos (n) y el número total de mujeres que recibieron el tratamiento (N). La siguiente columna, con el encabezamiento «Ctrl», corresponde a los datos del grupo control en cada estudio (aquellas mujeres que recibieron placebo o ningún tratamiento), el número de neonatos fallecidos (n) y el número total de mujeres en este grupo (N).

El cuadrado que aparece sobre cada línea del gráfico representa la *odds ratio* (la medida del efecto utilizada en este análisis) de cada estudio. En el caso de un resultado adverso como es la muerte, si el cuadrado se sitúa a la izquierda de la línea vertical del gráfico, indica que el tratamiento parece reducir el riesgo de muerte. Si se sitúa a la derecha, entonces indica que el tratamiento parece aumentar el riesgo de presentar el desenlace (muerte). El tamaño del cuadrado está en relación con la ponderación relativa de cada estudio particular con respecto al resultado combinado. La ponderación se basa en el número de participantes en cada estudio y el número de desenlaces (muertes). Los estudios más amplios y/o con mayor número de eventos tienen un mayor peso.

Las líneas horizontales representan el intervalo de confianza, es decir, el rango de valores dentro del cual se encontraría la *odds ratio* el 95% de las veces debido a la variación por el azar. Si la línea horizontal cruza la línea vertical central del gráfico, indica que no se puede establecer una conclusión firme a partir de ese estudio acerca de los beneficios o daños del tratamiento. Una línea cuyo extremo termina en forma de flecha indica que el intervalo de confianza se extiende más allá de la escala del gráfico.

La figura del diamante en la parte inferior del gráfico representa el resultado de la combinación de los datos de todos los estudios (metaanálisis). Su punto central representa la *odds ratio* combinada, y su amplitud representa el intervalo de confianza. El estadístico usado por defecto en la *Cochrane Library* es la *odds ratio* de Peto calculada según un modelo de efectos fijos. En este caso, la totalidad del diamante se halla a la izquierda de la línea central, lo que permite concluir con seguridad que el tratamiento es efectivo para reducir el riesgo de muerte de los neonatos, ya que este efecto no parece producirse por azar.

El resultado combinado muestra la etiqueta «Subtotal» en lugar de «Total» debido a que este gráfico representa un análisis de subgrupo dentro del análisis de la medida de resultado «Mortalidad neonatal» en esta amplia revisión.

En la parte más inferior del gráfico se muestra el valor de z, que representa el resultado de la prueba de significación del efecto global, es decir, una medida matemática equivalente a la posición y amplitud del diamante del gráfico.

El valor de chi-cuadrado representa la prueba de homogeneidad del tamaño del efecto observado a través de los diferentes estudios, es decir, mide hasta qué punto los resultados de los distintos estudios son concordantes entre sí.

ESTÁNDARES PARA LA PUBLICACIÓN DE REVISIONES SISTEMÁTICAS CON METAANÁLISIS DE ENSAYOS CLÍNICOS ALEATORIZADOS

Aunque, tal como se ha dicho, el número de revisiones sistemáticas y metaanálisis publicados ha aumentado de forma espectacular en los últimos años, no siempre se han explicitado suficientemente sus métodos, produciendo, por tanto, resultados poco claros o incluso contradictorios. Este hecho ha llevado a la necesidad de desarrollar normas para la publicación de este tipo de trabajos a fin de facilitar su interpretación y utilización por parte de los usuarios. Al igual que a mediados de los años noventa se publicó el informe CONSORT⁶ para la publicación de ensayos clínicos, en 1996 se creó el grupo QUOROM (*Quality of Reporting of Meta-analysis*), cuyo objetivo era establecer unas normas para mejorar la calidad de la presentación de los metaanálisis de ECA. Aunque previamente otros autores habían propuesto criterios y directrices para mejorar la calidad de la publicación de metaanálisis, QUOROM constituye la primera propuesta fruto de un amplio consenso que pretende ser de aplicación en las diversas áreas o disciplinas de las ciencias de la salud. QUOROM es un grupo de trabajo multidisciplinario, formado por epidemiólogos, clínicos, estadísticos, editores e investigadores. En 1996 tuvo lugar la Conferencia sobre Calidad de Elaboración de los Informes de los Metaanálisis, que dio como resultado la declaración QUOROM⁷, una lista de comprobación y un diagrama de flujo que facilitan la publicación de metaanálisis de buena calidad metodológica. Se puede consultar más información en: <http://www.consort-statement.org/QUOROM.pdf>.

DECLARACIÓN QUOROM

La propuesta QUOROM incluye una lista de comprobación (tabla 15-1) y un diagrama de flujo (fig. 15-2). La lista se estructura en seis categorías que se corresponden con cada una de las secciones que componen un artículo científico (Título, Resumen, Introducción, Métodos, Resultados y Discusión) y catorce subcategorías, relacionadas con aspectos más específicos que deberían incluirse en cada una de las categorías anteriores. En total, la lista contiene 18 ítems que deberán considerarse por parte de los autores de metaanálisis. La finalidad de esta lista de comprobación es animar a los autores a que proporcionen toda aquella información que resulta esencial para interpretar y utilizar adecuadamente los resultados de un metaanálisis.

A continuación se comentan brevemente cada uno de estos apartados.

Título

El título (ítem 1) del artículo debe identificar el trabajo como un metaanálisis o revisión sistemática de ECA. Un título informativo permite identificar más rápidamente un metaanálisis o una revisión sistemática de nuestro interés. Si

Tabla 15-1. Lista de comprobación en QUOROM

Categoría	Subcategoría	Descriptor	Notificado? (Sí/No)	N.º página
Título		Identificar el informe como un metaanálisis (o revisión sistemática de ECA)		
Resumen		Utilizar un formato estructurado <i>Describir:</i>		
	Objetivos	Explícitamente la pregunta clínica		
	Fuentes de datos	Las bases de datos (una lista) y otras fuentes de información		
	Métodos de revisión	Los criterios de selección (población, intervención, resultado y diseño de estudio) Los métodos para evaluar la validez, la extracción de datos y las características de los estudios, así como la síntesis cuantitativa suficientemente detallada para permitir su replicación		
	Resultados	Características de los ECA incluidos y excluidos, resultados cuantitativos y cualitativos (estimaciones puntuales e intervalos de confianza) y análisis de subgrupos		
	Conclusión	Los resultados principales		
Introducción		<i>Describir:</i> El problema clínico explícitamente, la justificación biológica para la intervención y la justificación de la revisión		
Métodos	Búsquedas	Las fuentes de información detalladas (como bases de datos, registros, ficheros personales, informes de expertos, agencias, búsqueda manual) así como cualquier restricción (años considerados, estado de la publicación, idioma de la publicación)		
	Selección	Los criterios de inclusión y exclusión (definiendo la población, la intervención, los resultados principales y el diseño de estudio)		
	Valoración de la validez	Criterios y proceso empleados (p. ej., condiciones de enmascaramiento, valoración de la calidad y sus hallazgos)		
	Extracción de datos	Proceso o procesos utilizados (p. ej., de forma independiente o por duplicado)		
	Características de los estudios	Tipo de diseño científico, características de los participantes, detalles de la intervención, variables de resultado y valoración de la heterogeneidad clínica		
	Síntesis cuantitativa de los datos	Las principales medidas de efecto (p. ej., el riesgo relativo), el método de combinación de los resultados (pruebas estadísticas e intervalos de confianza), manejo de los datos ausentes, valoración de la heterogeneidad estadística, justificación de cualquier análisis de sensibilidad o de subgrupo planificado a priori y cualquier valoración del sesgo de publicación		

Continúa

Tabla 15-1. Lista de comprobación en QUOROM

Categoría	Subcategoría	Descriptor	Notificado? (Sí/No)	N.º página
Resultados	Diagrama de flujo	Ilustrar la estructura del metaanálisis mediante el flujo de los ECA (v. fig. 2)		
	Características de los estudios	Presentar datos descriptivos para cada ensayo (p. ej. edad, tamaño de la muestra, intervención, dosis, duración, período de seguimiento)		
	Síntesis cuantitativa de los datos	Informar del grado de acuerdo en la valoración de la selección y la validez, presentar un resumen simple de los resultados (para cada grupo de tratamiento en cada ECA, para cada variable de resultado primaria), presentar los datos necesarios para calcular el tamaño de los efectos y los intervalos de confianza en los análisis por intención de tratar		
Discusión		Resumir los hallazgos clave, discutir las inferencias clínicas en función de la validez interna y externa, interpretar los resultados en función de la totalidad de la evidencia disponible, describir los potenciales sesgos en el proceso de revisión y sugerir áreas de investigación futura		

además especifica que se trata de una revisión sobre ECA, ayudará a priorizarla frente a otro tipo de revisiones.

Resumen

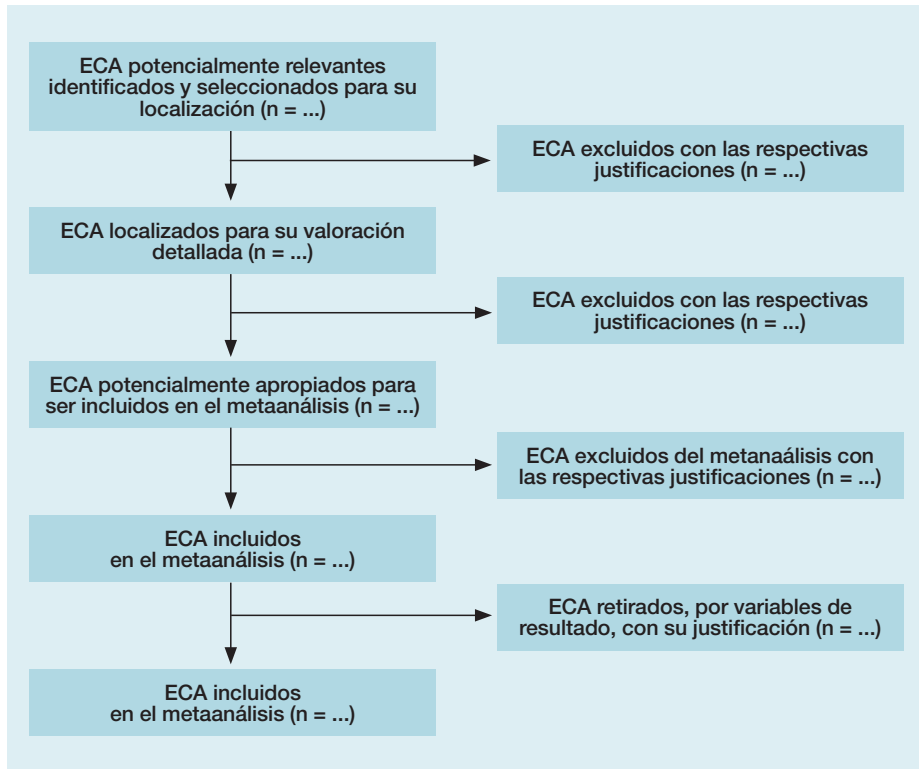
El resumen debe tener un formato estructurado (ítem 2), con una breve descripción de los siguientes apartados: objetivos (ítem 3), fuentes de datos (ítem 4), métodos de la revisión (ítem 5), resultados (ítem 6) y conclusión (ítem 7), los cuales se deberán describir con mayor detalle en la categoría correspondiente (v. más adelante). En los objetivos debe describirse de forma clara y precisa la pregunta clínica a la que la revisión desea dar respuesta. Como veremos más adelante, los objetivos se corresponden directamente con los criterios para la selección de los estudios de la revisión.

Introducción

En la introducción (ítem 8) debe describirse con detalle el problema clínico, los fundamentos biológicos de la intervención y la justificación de la revisión. Las revisiones sistemáticas no surgen de la nada, sino del contexto de un problema clínico relevante en el que existe incertidumbre sobre un determinado tratamiento o intervención. Por eso, el autor debe empezar por justificar la necesidad de la revisión haciendo referencia a la importancia sanitaria del problema y a la controversia existente sobre las intervenciones que se desean revisar.

Métodos

El apartado de métodos se compone de varios subapartados. En primer lugar, se solicita a los autores información exhaustiva sobre las *búsquedas bibliográficas* (ítem 9) llevadas a cabo para identificar los estudios: bases de datos, informes de agencias, guías de práctica clínica, así como registros de estudios, ficheros personales, informes de expertos, búsqueda manual, contactos con agencias de financiación de la investigación, industria farmacéutica, etc. Igualmente se solicita información sobre los años de publicación considerados, así como el estado de las publicaciones (sólo estudios publicados como originales en revistas o también en forma de resúmenes e incluso estudios no publicados) y el idioma de publicación. Existe controversia sobre la importancia de incluir datos no publicados en un metaanálisis⁸. Algunos estudios que han tratado de evaluar el papel de la «literatura gris» en los resultados de los metaanálisis encontraron que los metaanálisis que se limitaron a considerar sólo los ECA publicados tendían a sobreestimar el efecto de los tratamientos, en comparación con los que también incluyeron literatura gris. Sin embargo, otros estudios no mostraron resultados concluyentes en este sentido^{9,10}. Por lo que respecta al idioma de publicación, existe evidencia de que las restricciones idiomáticas en un metaanálisis pueden conducir a un resultado sesgado¹¹, especialmente en



© Elsevier. Fotocopiar sin autorización es un delito.

Figura 15-2. Diagrama de flujo de ensayos clínicos aleatorizados (ECA) en QUOROM.

algunos ámbitos, por lo que se anima a los autores a no restringir las búsquedas a un único idioma, habitualmente el inglés.

A continuación se solicitan datos sobre la *selección* (ítem 10) de los estudios, es decir, los criterios de inclusión y exclusión. La selección de estudios para una revisión no puede depender en ningún caso de la dirección de sus resultados, sino que deben establecerse *a priori* unos criterios explícitos que guíen esta importante decisión por parte de los autores. Estos criterios, que se desprenden directamente de la pregunta formulada en los objetivos del trabajo, consisten en definir con suficiente detalle y precisión la población de interés (en otras palabras, el problema clínico o condición), la intervención (incluye la necesidad de establecer también el grupo control de comparación), las principales variables de resultado que se consideran relevantes (a menudo se requiere de su definición u operativización) y el diseño de estudio que se considera adecuado para la revisión. De estos criterios no sólo dependerán las decisiones acerca de la selección de estudios, sino también el diseño de la estrategia de búsqueda.

El siguiente apartado hace referencia a la *valoración de la calidad* (ítem 11) de los estudios incluidos. Existe evidencia directa¹²⁻¹⁵ de la que se deriva la necesidad de documentar la calidad de los ECA incluidos en el metaanálisis, ya que numerosos estudios han demostrado que la inclusión de estudios de baja calidad tiende a que se sobrestimen los efectos de la intervención valorada. Por ello, es importante que el lector de la revisión pueda conocer la calidad de las pruebas a partir de las cuales se construye la misma y se generan sus conclusiones. Es por este motivo que QUOROM establece la necesidad de explicitar los criterios (p. ej., asignación aleatoria, encubrimiento de las asignaciones, enmascaramiento y pérdidas en el seguimiento) y el proceso utilizado (p. ej., utilización de escalas y *scores*). Tratándose de un aspecto cualitativo de tanta importancia que está sujeto a un juicio subjetivo, es relevante conocer quién o quiénes aplicaron estos criterios y cómo. Los expertos, entre ellos la Colaboración Cochrane, sugieren que se haga al menos por pares de forma independiente, previendo un sistema para resolver las posibles discrepancias.

Posteriormente, debe presentarse el proceso o procesos empleados para la *extracción de los datos* (ítem 12) de los estudios (p. ej., de forma independiente, por duplicado, utilizando una hoja de extracción de datos diseñada de forma específica para la revisión) y deben describirse las *características de los estudios* (ítem 13): tipo de diseño, características de los participantes, detalles de la intervención, definición de las variables de resultado y valoración de la heterogeneidad clínica. Habitualmente se recomienda acompañar esta sección de la revisión con una tabla-resumen.

El último apartado de métodos hace referencia a la *síntesis cuantitativa de los datos* (ítem 14). Deben especificarse las principales medidas de efecto (como el riesgo relativo), los métodos estadísticos empleados para la combinación de los resultados (ya que existen diversas estrategias de análisis alternativas, unas más conservadoras que otras, que deben explicitarse), el manejo de los datos no disponibles (debe conocerse si se imputaron o no valores para los datos no disponibles en el cálculo de las medidas de efecto) y la valoración de la heterogeneidad estadística entre los estudios incluidos (sólo está justificado

aplicar técnicas de metaanálisis cuando los resultados de los diferentes estudios que se pretende combinar son razonablemente homogéneos o consistentes). Asimismo, debe justificarse *a priori* cualquier análisis de sensibilidad y de subgrupo. Finalmente, se recomienda documentar la estrategia utilizada para analizar el sesgo de publicación.

Resultados

En el apartado de resultados debe presentarse el *diagrama de flujo de ensayos clínicos* (ítem 15) (v. fig. 15-2), en el que se describe todo el proceso desde la identificación inicial de los ECA potencialmente relevantes hasta la selección definitiva de los mismos. A su vez, deben detallarse los eventuales motivos para la exclusión de cualquier trabajo, pues es importante que el lector perciba que la selección de estudios en la revisión no ha sido un proceso arbitrario. A continuación, deben presentarse las *características de cada uno de los estudios* (ítem 16), como la edad de los pacientes, el tamaño de la muestra, la intervención, la dosis, la duración y el período de seguimiento). Esta información permitirá al lector conocer hasta qué punto los estudios finalmente incluidos en la revisión sistemática son homogéneos desde el punto de vista clínico, lo que a su vez ayudará a valorar la pertinencia o no de un metaanálisis y su aplicabilidad.

El último apartado de los resultados lo constituye la *síntesis cuantitativa de los datos* (ítem 17). En primer lugar, debe presentarse un informe sobre el grado de acuerdo entre los autores de la revisión sobre la selección de los estudios y la valoración de su calidad. Un alto grado de acuerdo garantiza que los criterios se han aplicado de forma consistente. Posteriormente, deben exponerse de forma clara y ordenada los resultados para cada grupo de tratamiento en cada ECA y para cada variable de resultado relevante. Además, hay que proporcionar los datos necesarios para calcular el tamaño del efecto y los intervalos de confianza según un análisis por intención de tratar, por ejemplo mediante tablas de 2×2 , medias y desviaciones estándar, proporciones, etc. Sólo una presentación clara de los datos permite al lector valorar e interpretar adecuadamente los resultados. En ocasiones, el lector puede desear reanalizar los datos según un criterio distinto al de la revisión, para lo que necesitará disponer de los datos crudos, completos y presentados de forma ordenada.

Discusión

Finalmente, en la discusión (ítem 18) deben resumirse los resultados principales, discutiendo las implicaciones para la práctica clínica a partir de la validez interna y externa de la revisión. Los resultados tienen que interpretarse en función de la totalidad de la evidencia disponible. También deben describirse los sesgos potenciales que han podido incidir en el proceso de revisión y que podrían cuestionar las conclusiones o imponer mucha cautela en su interpretación. Finalmente, puesto que las revisiones sistemáticas también permiten delimitar las áreas de incertidumbre en las que no existe evidencia o ésta es controvertida, permiten señalar futuras líneas de investigación necesarias para cubrir estas lagunas en el conocimiento.

IMPACTO DE QUOROM EN LA CALIDAD DE LAS PUBLICACIONES

Algunos aspectos criticados de la versión inicial de QUOROM es que no especifica la afiliación o fuentes de financiación de los autores o la comparabilidad entre los grupos de tratamiento de los documentos originales¹⁶. Estas normas de publicación están siendo actualmente revisadas por sus autores y cabe prever una nueva versión modificada en los próximos meses, que adoptará el nuevo nombre de PRISMA (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*)¹⁷.

Aunque la aparición de las iniciativas mencionadas refleja la necesidad de mejorar la publicación de todo tipo de estudios en las revistas biomédicas, sólo CONSORT parece haber logrado un alto grado de aceptación, por lo que ha sido incorporada por la *World Association of Medical Editors*, el *International Committee of Medical Journals Editors* (ICMJE) y el *Council of Science Editors*¹⁸. No ha sido ésta todavía la suerte de QUOROM, pues no se han realizado estudios sobre su impacto en cuanto a la mejoría de la calidad de los metaanálisis después de su publicación ni de su inclusión como criterio editorial de las revistas biomédicas. Aunque la aplicación del *checklist* de QUOROM no resulta complejo ni consume excesivo tiempo, desafortunadamente muchas revistas, en especial las de especialidades médicas donde se publica el mayor número de las revisiones sistemáticas, todavía no lo han adoptado, por lo que se plantean dudas acerca de la fiabilidad de sus resultados y conclusiones.

Algunos estudios han tratado de evaluar el grado de cumplimiento de las recomendaciones de QUOROM y su impacto a lo largo del tiempo. En un estudio reciente¹⁹ se identificaron todos los metaanálisis de ECA (161) publicados en Medline entre los años 2000 y 2005, y se aplicaron los criterios de QUOROM. La puntuación media obtenida fue de 12,3 puntos (sobre un máximo de 18), con un rango entre 4,5 y 16,5. Lo más significativo fue la constatación de un incremento en la puntuación media a lo largo del tiempo (10,5 en 2000 hasta 13,0 en 2005). Se halló que las revisiones Cochrane tenían mayor calidad que los metaanálisis publicados en revistas. Aunque se observó una mejoría a lo largo del tiempo en la mayoría de ítems de QUOROM, no obstante, persistieron las insuficiencias en algunos de ellos, como son la comunicación de los resultados (ítem 17), la discusión (ítem 18) y, sobre todo, el diagrama de flujo de los estudios (ítem 15).

Otro estudio reciente²⁰ identificó 300 revisiones sistemáticas indexadas en PubMed en noviembre de 2004 y publicadas en inglés. En su mayoría consistían en revisiones terapéuticas (71%) publicadas en revistas de especialidades médicas (91%). En el 54% se había realizado un metaanálisis. Nuevamente, se halló que las revisiones Cochrane tenían mayor calidad que las revisiones publicadas en revistas biomédicas con sistema de revisión por pares. Algunas de las explicaciones para esta superioridad de las revisiones Cochrane son la existencia de un protocolo previo a la revisión y el compromiso de realizar actualizaciones periódicas. El proceso de elaboración y publicación de una revisión Cochrane es un proceso altamente estructurado y estandarizado tal como queda reflejado en el Manual de Revisores Cochrane (disponible *on line* en: <http://>

www.cochrane-handbook.org). Entre otros aspectos, este manual establece la estructura de una revisión sistemática Cochrane, tal como queda esquematizado en la tabla 15-2.

El esquema de una revisión Cochrane no es exactamente igual al que propone QUOROM, pues la publicación en la *Cochrane Library* (y en la Biblioteca Cochrane Plus, la versión en castellano) no tiene las limitaciones de espacio habituales de las revisiones que se publican en revistas biomédicas, por lo que

Tabla 15-2. Esquema de una revisión Cochrane

Carátula

- Título
- Revisores
- Fuentes de apoyo
- Qué contenido es nuevo

Texto de la revisión

- Sinopsis
- Resumen
- Antecedentes
- Objetivos
- Criterios para la selección de los estudios de esta revisión
 - Tipos de estudios
 - Tipos de participantes
 - Tipos de intervenciones
 - Tipos de medidas de resultado (outcomes)
- Estrategia de búsqueda para la identificación de los estudios
- Métodos de revisión
- Descripción de los estudios
- Calidad metodológica
- Resultados
- Discusión
- Conclusiones de los revisores
 - Implicaciones para la práctica
 - Implicaciones para la investigación
- Agradecimientos
- Conflictos de interés

Referencias

- Referencias de los estudios
 - Estudios incluidos
 - Estudios excluidos
 - Estudios pendientes de evaluación
 - Estudios en curso
- Otras referencias
 - Referencias adicionales
 - Otras versiones publicadas de esta revisión

Tablas y figuras

- Características de los estudios incluidos
- Características de los estudios excluidos
- Características de los estudios en curso
- Comparaciones, datos y gráficos
- Tablas adicionales

suelen ser mucho más exhaustivas. No obstante, ambos formatos comparten unos mismos criterios de rigor, transparencia y precisión y las iniciativas que las impulsan reflejan una misma preocupación y un mismo compromiso en garantizar la calidad y utilidad para el lector de las revisiones sistemáticas, a fin de que sean instrumentos válidos para la toma de decisiones.

Además, el estudio halló importantes déficits en un porcentaje elevado de las revisiones publicadas en revistas biomédicas, como, por ejemplo, la no especificación del período de la búsqueda bibliográfica, la falta de información sobre la calidad de los estudios, la ausencia de un análisis del sesgo de publicación, la falta de detalles acerca de los criterios de elegibilidad, o bien la ausencia de valoración de los potenciales riesgos de la intervención. También existían sospechas de un sesgo por la notificación selectiva de resultados. Aunque la limitación de espacio en las revistas biomédicas puede suponer una dificultad que contribuye a que la información sea incompleta, la falta de unos criterios claros por parte de los autores parece ser la razón fundamental de este problema.

Estas razones apoyan la necesidad de un uso extendido de QUOROM y su adopción por parte de los editores de las revistas biomédicas, dada la importancia creciente de las revisiones en la literatura biomédica y como herramienta de soporte en la toma de decisiones. Esta aspiración se ve reforzada por el hecho de que la calidad de las revisiones publicadas en revistas que han adoptado los criterios QUOROM también es mayor que la de aquellas que no lo han hecho²¹.

ESTÁNDARES PARA LA PUBLICACIÓN DE METAANÁLISIS DE ESTUDIOS OBSERVACIONALES EN EPIDEMIOLOGÍA

Existen muchas situaciones donde el ensayo clínico controlado y aleatorizado no es posible, y sólo se dispone de datos procedentes de estudios observacionales. Por ejemplo, en los estudios acerca de factores de riesgo o de exposiciones que pueden causar daño, la exposición no puede asignarse al azar, bien porque está relacionada con características inherentes a los sujetos de estudio (p. ej., grupo sanguíneo, sexo o una característica genética determinada), bien porque no resulta ético ni aceptable exponer a los individuos a exposiciones que podrían causarles daño (p. ej., al asbesto o una radiación ionizante). Los estudios observacionales también se han utilizado habitualmente para el estudio de la efectividad de intervenciones complejas en el campo de la salud pública, como programas sanitarios o intervenciones educativas, dadas las dificultades prácticas para realizar un estudio experimental en este contexto o también la necesidad de explorar el efecto de la intervención en la comunidad en condiciones normales.

A diferencia de los estudios experimentales, en los estudios observacionales la asignación a la exposición (estudios etiológicos) o intervención (estudios de efectividad) de interés no depende del azar, y se estudia la asociación entre los cambios o diferencias observadas en una característica (la exposición o la intervención) y los cambios o diferencias observados en una medida de resultado de

interés. Algunos diseños de estudios observacionales son los que utilizan datos procedentes de bases de datos existentes, los estudios de corte transversal, las series de casos, los estudios de casos y controles, los estudios con controles históricos y los estudios de cohortes.

La aplicación del metaanálisis en los estudios observacionales ha sido un tema muy controvertido, dado el mayor riesgo potencial de sesgo en este tipo de estudios en comparación con los ensayos clínicos controlados y aleatorizados. Además, la gran diversidad de diseños en los estudios observacionales y en las poblaciones estudiadas hace particularmente problemática la interpretación de una medida resumen a partir de la combinación de sus resultados. Por esta razón, no es de extrañar que exista un claro predominio de metaanálisis de ensayos clínicos en la literatura biomédica. No obstante, el número de metaanálisis de estudios observacionales publicados ha aumentado progresivamente, pues son una herramienta de gran ayuda en la toma de decisiones especialmente en el área de la salud pública y la política sanitaria. Por esta razón, y dados los retos particulares que presentan este tipo de estudios a causa del riesgo inherente de sesgos y las diferencias en los diseños de los estudios observacionales, se hacía necesario disponer de una herramienta que ayudara a mejorar la calidad de los informes publicados de este tipo de metaanálisis.

En 1997 se creó el grupo MOOSE (*The Meta-analysis of Observational Studies in Epidemiology*), formado por 27 expertos procedentes de diferentes áreas (clínicos, investigadores clínicos y sociales, estadísticos, epidemiólogos y editores), cuyo objetivo era confeccionar una lista de comprobación que ayudara a mejorar la calidad de la presentación de los metaanálisis de estudios observacionales. El grupo de trabajo celebró un seminario en 1997 que dio como resultado la propuesta MOOSE²².

PROPUESTA MOOSE

La propuesta MOOSE consiste en una lista de comprobación (tabla 15-3) donde se establece un total de 35 recomendaciones estructuradas en seis grandes apartados: antecedentes, estrategia de búsqueda, métodos, resultados, discusión y conclusiones.

A continuación se comentan brevemente cada uno de estos apartados.

Introducción

La introducción debe incluir una definición del problema, la hipótesis del estudio, una descripción de la(s) medida(s) de resultado considerada(s), del tipo de exposición o intervención estudiado, del tipo de diseño utilizado, así como una descripción detallada de la población de estudio. Esta descripción detallada resulta especialmente pertinente en este tipo de metaanálisis donde cabe esperar un elevado grado de heterogeneidad entre los estudios seleccionados (diferencias en la población de estudio, en el diseño y/o en las medidas de resultado) que repercutirá en los resultados.

Tabla 15-3. Lista de comprobación MOOSE (propuesta para autores, editores y revisores de metaanálisis de estudios observacionales)

- La sección de los *antecedentes* debería incluir:
 - Definición del problema
 - Hipótesis planteada
 - Descripción de la(s) medida(s) de resultado
 - Tipo de exposición o intervención utilizada
 - Tipo de diseños de estudio utilizados
 - Población de estudio
- La sección de la *estrategia de búsqueda* debería incluir:
 - Calificación de quien realizó la búsqueda (documentalistas e investigadores)
 - Estrategia de búsqueda, incluyendo el período de tiempo y palabras clave utilizadas
 - Esfuerzos realizados para incluir todos los estudios disponibles, incluyendo contacto con los autores
 - Bases de datos y registros consultados
 - *Software* utilizado para la búsqueda, su nombre y versión, incluyendo funcionalidades especiales (p. ej., explosión)
 - Búsqueda manual (p. ej., lista de referencias de los estudios recuperados)
 - Listado de las citaciones localizadas y de las excluidas, con su justificación
 - Método empleado para los estudios publicados en otros idiomas distintos del inglés
 - Método empleado para los estudios no publicados y los resúmenes
 - Descripción de los contactos con autores
- La sección de los *métodos* debería incluir:
 - Descripción de la relevancia y justificación de los estudios seleccionados para demostrar la hipótesis del estudio
 - Justificación de la selección y sistema de codificación de los datos (p. ej., según criterios clínicos establecidos o por conveniencia)
 - Documentación acerca de cómo los datos fueron clasificados y codificados (p. ej., múltiples evaluadores, cegamiento y fiabilidad interobservador)
 - Valoración de los factores de confusión (p. ej., comparabilidad entre los casos y controles)
 - Valoración de la calidad de los estudios, incluyendo el cegamiento de los evaluadores; estratificación o regresión para posibles factores predictivos de los resultados
 - Valoración de la heterogeneidad
 - Descripción suficientemente detallada de los métodos estadísticos (p. ej., una descripción completa del modelo de efectos fijos o aleatorios, justificación acerca de si el modelo seleccionado tiene en cuenta los factores predictores de los resultados, modelos de dosis-respuesta, o metaanálisis acumulativo) para poder ser reproducidos
 - Presentación adecuada de los resultados en forma gráfica o en tablas
- La sección de los *resultados* debería incluir:
 - Resumen gráfico de los resultados (estimadores) individuales y globales
 - Tabla descriptiva con información de todos los estudios incluidos
 - Resultados de los análisis de sensibilidad o de subgrupo
 - Medida de la incertidumbre estadística de los resultados
- La sección de la *discusión* debería incluir:
 - Valoración cuantitativa de los sesgos (p. ej., sesgo de publicación)
 - Justificación de las exclusiones (p. ej., exclusión de estudios en idioma distinto al inglés)
 - Valoración de la calidad de los estudios incluidos
- La sección de las *conclusiones* debería incluir:
 - Posibles explicaciones alternativas para los resultados observados
 - Generalización de las conclusiones (p. ej., a partir de los datos disponibles y en el contexto de lo que se sabe sobre el tema)
 - Orientaciones para la investigación futura
 - Fuentes de financiamiento

Estrategia de búsqueda

La estrategia de búsqueda debe especificar la cualificación de quién realiza la búsqueda, las bases de datos electrónicas y otros registros utilizados, el diseño

de la estrategia de búsqueda con los términos empleados y el período, la utilización de *software* específico, la búsqueda manual y el contacto con autores, las restricciones idiomáticas en la selección de los estudios, el uso de material no publicado y de resúmenes, así como los criterios de exclusión. Algunos estudios han puesto de manifiesto las insuficiencias en la descripción de la estrategia de búsqueda en la mayoría de metaanálisis publicados²³.

Métodos

Los métodos deben incluir una descripción de la relevancia de los estudios seleccionados y justificación de su combinación en el metaanálisis, el grado en que la clasificación y la codificación de los datos en los estudios estaban especificadas y fueron objetivas, la valoración de los factores de confusión, de la calidad de los estudios y de la heterogeneidad entre éstos, una descripción de los métodos estadísticos utilizados en el análisis y la presentación de los resultados en forma gráfica o en tablas. La valoración de la calidad de los estudios resulta de particular importancia por su potencial efecto sobre los resultados. Por ello, resulta muy recomendable explorar, mediante análisis de subgrupo o de sensibilidad, cómo determinadas características clave del diseño están influyendo en los resultados y cómo ello explica la variabilidad observada entre los estudios. Los análisis estratificados o de regresión pueden también ser muy útiles para este propósito. En resumen, las fuentes de sesgo y de heterogeneidad pueden ser previstas *a priori* y, en consecuencia, confirmadas en el análisis.

Resultados

Los resultados deben incluir un resumen gráfico de los resultados (estimaciones) de los estudios incluidos en la revisión y de cualquier combinación de los mismos, una tabla descriptiva con información esencial de cada estudio, resultados de los análisis de sensibilidad o de cualquier análisis de subgrupo realizado, y una medida de la incertidumbre estadística de los resultados.

Discusión

La discusión debe abordar el problema de los sesgos, incluido el sesgo de publicación, los factores de confusión y la calidad de los estudios. Tanto los defectos en el diseño de los estudios, como en el modo en que los estudios son seleccionados en la revisión, tienden a distorsionar la magnitud o la dirección de las asociaciones observadas y, por tanto, cuestionan la validez del metaanálisis. Las variaciones en la calidad de los estudios pueden explicar a su vez la variabilidad en sus resultados.

Conclusión

Debido al elevado riesgo de sesgo inherente a los estudios observacionales, la conclusión debe tener en cuenta otras posibles explicaciones alternativas para los

resultados observados, así como el grado en que la generalización de éstos resulta o no adecuada (a partir de los datos disponibles y en el contexto de lo que se sabe sobre el tema). El metaanálisis también puede ayudar a identificar áreas que requieren mayor investigación. Las fuentes de financiamiento deben ser reveladas y su influencia, explorada en los resultados en caso de ser necesario.

BIBLIOGRAFÍA

1. Ceballos C, Valdizan JR, Artal A, Almarcegui, Allepuz C, García Campayo, et al. ¿Por qué la medicina basada en la evidencia? 20 años de metaanálisis. *An Med Interna*. 2000;17:521-6.
2. Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and reporting characteristics of systematic reviews. *PLoS Med*. 2007;4:e78.
3. Gispert JP, Bonfill X. ¿Cómo realizar, evaluar y utilizar revisiones sistemáticas y metaanálisis? *Gastroenterol Hepatol*. 2004;27:129-49.
4. Alderson P, Green S, Higgins JPT, eds. *Cochrane Reviewers' Handbook 4.2.2* [updated December 2003]. en: *The Cochrane Library, Issue 1, 2004*. Chichester, Reino Unido: John Wiley & Sons, Ltd.; 2003.
5. <http://www.cochrane-handbook.org/>.
6. Begg CB, Cho MK, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials: the CONSORT statement. *JAMA*. 1996;276:637-9.
7. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF, et al. Improving the quality of reports of meta-analyses of randomised controlled trials: The QUOROM statement. *Lancet*. 1999;354:1896-900.
8. Cook DJ, Guyatt GH, Ryan G, Clifton J, Buckingham L, Willan A, et al. Should unpublished data be included in meta-analysis? Current convictions and controversies. *JAMA*. 1993;269:2749-53.
9. Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet*. 1991;337:867-72.
10. Smith ML. Publication bias and meta-analysis. *Eval Educ*. 1980;4:22-4.
11. Moher D, Pham B, Klassen TP, et al. Does the language of publication of reports of randomized controlled trials influence the estimates of intervention effectiveness reported in meta-analyses? 6th Cochrane Colloquium; 1998.
12. Schultz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA*. 1995;273:408-12.
13. Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, et al. Does the quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet*. 1998;352:609-13.
14. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: Is blinding necessary? *Control Clin Trials*. 1996;17:1-12.
15. Berlin JA, on behalf of the University of Pennsylvania meta-analysis blinding study group. Does blinding of readers affect the results of meta-analyses? *Lancet*. 1997;350:185-6.
16. Christensen E. Quality of reporting of meta-analyses: The QUOROM statement. Will it help? *J Hepatol*. 2001;34:342-5.
17. The PLoS Medicine Editors. Many reviews are systematic but some are more transparent and completely reported than others. *PloS Med*. 2007;4:e147.
18. Altman DG. Endorsement of the CONSORT statement by high impact medical journals: Survey of instructions for authors. *BMJ*. 2005;330:1056-7.

19. Wen J, Ren Y, Wang L, Li Y, Liu Y, Zhou M, et al. The reporting quality of meta-analyses improves: A random sampling study. *J Clin Epidemiol*. 2008;61:770-5.
20. Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and reporting characteristics of systematic reviews. *PLoS Med*. 2007;4:e78.
21. Mrkobrada M, Thiessen-Philbrook H, Haynes B, Iansavichus AV, Rehman F, Garg AX. Need for quality improvement in renal systematic reviews. *Clin J Am Soc Nephrol*. 2008;3:1102-14.
22. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology. A proposal for reporting. *JAMA*. 2000;283:2008-12.
23. Sipe TA, Curlette WL. A meta-synthesis of factors related to educational achievement. *Int J Educ Res*. 1997;25:583-98.





Artículo especial

Declaración PRISMA: una propuesta para mejorar la publicación de revisiones sistemáticas y metaanálisis

PRISMA declaration: A proposal to improve the publication of systematic reviews and meta-analyses

Gerard Urrútia^{a,b,*} y Xavier Bonfill^{a,b,c}

^a Servei d'Epidemiologia Clínica i Salut Pública, Centro Cochrane Iberoamericano, Hospital de la Sant Creu i Sant Pau, Barcelona, España

^b CIBER Epidemiología y Salud Pública, Barcelona, España

^c Medicina Preventiva, Departamento de Pediatría, Obstetricia y Ginecología, Universitat Autònoma de Barcelona, Barcelona, España

INFORMACIÓN DEL ARTÍCULO

Historia del artículo:

Recibido el 3 de diciembre de 2009

Aceptado el 12 de enero de 2010

On-line el 5 de marzo de 2010

Precedente de la declaración QUOROM

En el año 1999, después de 3 años de trabajo, se publicó la declaración QUOROM, cuyo objetivo era establecer unas normas para mejorar la calidad de la presentación de los metaanálisis de ensayos clínicos aleatorizados¹. Se publicó un comentario de este documento en un monográfico sobre listas de comprobación para autores, revisores y editores de revistas médicas en Medicina Clínica².

En resumen, la declaración Quality Of Reporting Of Meta-analysis (QUOROM) incluye una lista de comprobación estructurada con 18 ítems que los autores de un metaanálisis, y también los editores de revistas, deberían considerar a la hora de publicar su trabajo en forma de artículo en una revista médica. Además, incluye un diagrama de flujo que describe todo el proceso, desde la identificación inicial de los estudios potencialmente relevantes hasta la selección definitiva de éstos. La finalidad de QUOROM era animar a los autores a que proporcionaran toda aquella información que resulta esencial para interpretar y utilizar adecuadamente los resultados de un metaanálisis.

Numerosos estudios realizados con posterioridad a la publicación de QUOROM han mostrado que la calidad de los metaanálisis publicados en revistas médicas todavía es deficiente. A pesar de esto, y a diferencia de otras iniciativas similares como CONSORT (dirigida a ensayos clínicos), la declaración QUOROM no parece

haber logrado el mismo grado de aceptación por parte de los editores de revistas biomédicas, aun cuando el número de revisiones sistemáticas y metaanálisis que se publican anualmente es muy elevado y en número creciente (se estima en 2.500 solamente las revisiones publicadas en inglés)³. Una de las causas de este menor éxito de QUOROM podría ser que se han realizado pocos estudios que demuestren su impacto en la mejoría de la calidad de los metaanálisis después de su publicación o de su inclusión como criterio editorial de las revistas biomédicas⁴.

En el momento de su publicación, el grupo de trabajo QUOROM estableció la necesidad de una revisión y una actualización periódica de las directrices conforme a la nueva evidencia publicada, y que podía comportar mantener, eliminar o añadir nuevos ítems a la lista de comprobación inicial. Como resultado, en julio de 2009 se publicó la declaración PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses), una actualización y ampliación de QUOROM.

Declaración PRISMA

A diferencia de QUOROM, la declaración PRISMA⁵ viene acompañada de un extenso documento donde se detalla la explicación o la justificación de cada uno de los 27 ítems propuestos, así como el proceso de elaboración de estas directrices⁶. Reflejo del objetivo pedagógico de este documento, se incluyen 7 cuadros que proporcionan una explicación más detallada de ciertos aspectos clave sobre la metodología y la conducción de revisiones sistemáticas (terminología, formulación de la pregunta de investigación, identificación de los estudios y extracción de datos, calidad de los estudios y riesgo de sesgo,

* Autor para correspondencia.

Correo electrónico: gurrutia@santpau.cat (G. Urrútia).

Este autor está adscrito al Programa de Doctorado en Salud Pública y Metodología de la Investigación Biomédica de la Universitat Autònoma de Barcelona.

cuándo combinar datos, metaanálisis y análisis de la consistencia, y sesgo de publicación selectiva de estudios o resultados).

Como señalan sus autores, PRISMA incorpora varios aspectos conceptuales y metodológicos novedosos relacionados con la metodología de las revisiones sistemáticas que han emergido en los últimos años, período en el que ha habido una importante producción de revisiones y de investigación sobre éstas. Uno de ellos es el uso de la terminología utilizada para describir una revisión sistemática y un metaanálisis, hasta la fecha algo confusa e inconsistente. Los autores de PRISMA han adoptado las definiciones de la Colaboración Cochrane⁷. Según ésta, el metaanálisis (síntesis cuantitativa de resultados) sería solamente una parte, deseable pero no siempre posible, de un proceso más amplio —consistente en diversos pasos sucesivos que deben ser explícitos y reproducibles—, conocido como revisión sistemática. Por otro lado, PRISMA tiene una aplicabilidad más amplia que su predecesor QUOROM, ya que no se limita solamente a los metaanálisis de ensayos clínicos aleatorizados, sino que también es útil para las revisiones de otro tipo de estudios. La primera consecuencia de estos cambios se refleja en el nombre de PRISMA: «...directrices para la publicación de revisiones sistemáticas y metaanálisis de estudios que evalúan intervenciones sanitarias».

Cabe señalar que la elaboración y la publicación de PRISMA ha coincidido en el tiempo con la actualización y la modificación sustancial del Cochrane Handbook for Systematic Reviews of Interventions, versión 5 (Manual del Revisor Cochrane), cuya finalidad principal es ayudar a los autores de revisiones Cochrane a ser sistemáticos y explícitos en el desarrollo de éstas⁸. No en vano muchos de los 29 autores del grupo de trabajo PRISMA están también involucrados como asesores metodológicos de la Colaboración Cochrane. Por esta razón, el Manual del Revisor Cochrane incorpora muchos de los mismos cambios propuestos en PRISMA.

Aspectos conceptuales novedosos introducidos en PRISMA

Los autores de PRISMA identifican 4 aspectos conceptuales novedosos que conllevan la adición de nuevos ítems a la lista de comprobación:

1. El carácter iterativo del proceso de desarrollo de una revisión sistemática. La conducción de una revisión sistemática es un proceso complejo que implica numerosos juicios y decisiones por parte de los autores. Con el fin de minimizar el riesgo de sesgo en el proceso de la revisión, estos juicios y decisiones no deberían estar influidos por los resultados de los estudios incluidos en la revisión. El conocimiento anticipado que los autores puedan tener de los resultados de los estudios potencialmente elegibles podría, por ejemplo, influir en la pregunta que la revisión trata de responder, en los criterios de selección de los estudios, en la elección de las comparaciones que se van a analizar o en los resultados que se van a reportar en la revisión. Dada la naturaleza retrospectiva de las revisiones, es muy importante que los métodos que utilice la revisión se establezcan y se documenten a priori. La publicación previa del protocolo, tal como ocurre en las revisiones Cochrane, reduce el impacto de los sesgos inherentes al autor y promueve la transparencia acerca de los métodos y del proceso, además de evitar revisiones redundantes. La existencia de un protocolo no excluye que pueda haber razones justificadas para modificar el protocolo de revisión original. En este caso, es importante garantizar que los cambios (por ejemplo, la exclusión de estudios previamente seleccionados) no se han tomado a posteriori al saber cómo afectarán a los resultados. Tales decisiones son muy susceptibles de introducir sesgos y deben evitarse.

Por todas las razones hasta aquí apuntadas, es importante que se detalle si existía un protocolo previo a la revisión y que se hagan explícitos los cambios introducidos, así como su justificación. Solo así es posible juzgar si tales cambios a posteriori fueron apropiados o no y si podrían introducir algún tipo de sesgo.

2. La conducción y la publicación de un estudio de investigación son conceptos distintos. Aunque esta distinción es menos evidente para una revisión sistemática que en el caso de un estudio primario, la publicación y la conducción de una revisión sistemática están muy entrelazadas. Por ejemplo, no comunicar si se ha realizado o no la evaluación del riesgo de sesgo de los estudios incluidos en una revisión es señal de una conducción deficiente dada la importancia de este aspecto en el proceso de la revisión.
3. Evaluación del riesgo de sesgo al nivel de los estudios o de los resultados. El grado en que una revisión puede arrojar conclusiones fiables sobre los efectos de una intervención depende de la validez de los datos y los resultados de los estudios incluidos en la revisión. Así, por ejemplo, un metaanálisis de estudios con poca validez interna producirá unos resultados engañosos. Por esto, la evaluación de la validez de los estudios incluidos es un componente esencial de una revisión y debe considerarse en los análisis, la interpretación y las conclusiones de ésta. La declaración QUOROM, y también las versiones previas del Manual del Revisor Cochrane, se referían a este importante aspecto como a la «evaluación de la calidad». No obstante, hoy se recomienda evitar el término «calidad» para hablar de «riesgo de sesgo», ya que el primero se refiere a la valoración de hasta qué punto el estudio se ha llevado a cabo según los mejores estándares posibles, lo que no implica necesariamente que se haya evitado el riesgo de sesgo en los resultados. Por ejemplo, el mejor estudio posible bajo determinadas circunstancias puede tener un importante riesgo de sesgo. El aspecto clave en una revisión es preguntarse hasta qué punto los resultados de los estudios incluidos pueden tener credibilidad, de modo que la evaluación del «riesgo de sesgo» aborda esta misma cuestión. Además, la valoración de la calidad del estudio está condicionada por la calidad del reporte y la disponibilidad de información detallada acerca de los métodos, información de la que no siempre se dispone. De entre las numerosas opciones existentes para valorar el riesgo de sesgo, los autores de PRISMA promueven un sistema basado en la evaluación de diversos componentes clave del diseño y la ejecución de los estudios para los que existen sólidas evidencias empíricas acerca de su relación con el sesgo, como es el caso del nuevo sistema adoptado por la Colaboración Cochrane⁹. La evaluación rigurosa del riesgo de sesgo requiere tanto de una evaluación al nivel de los estudios (por ejemplo, evaluar aspectos genéricos del diseño, como la ocultación de la secuencia aleatorizada) como, en algunas ocasiones, también de los resultados (fiabilidad y validez de los datos para cada resultado específico a partir de los métodos utilizados para su medición en cada estudio individual). La calidad de la evidencia puede diferir entre resultados, aun dentro de un mismo estudio. Por ejemplo, algunas variables de resultado pueden haberse registrado de manera muy cuidadosa y exhaustiva, mientras que otras no tanto. Esta información debe referirse para permitir una evaluación explícita de hasta qué punto la estimación del efecto es correcta.
4. Importancia de los sesgos relacionados con la publicación. Las revisiones sistemáticas deben tratar de incorporar información de todos los estudios que sean relevantes para el tema de la revisión. No obstante, la ausencia de información de algunos estudios puede cuestionar la validez de la revisión. Este problema puede ocurrir cuando no se publican estudios completos (sesgo de publicación) o porque la información

publicada es incompleta o inexacta (por ejemplo, el informe selectivo de resultados). Tanto en un caso (omisión de estudios completos, habitualmente por causa de unos resultados no favorables) como en otro (omisión de resultados dentro de un estudio individual) existe evidencia empírica que demuestra su relación con el sesgo al estar basada la revisión en una muestra sesgada y no representativa de toda la información existente (publicada o no). Por todas estas razones, los autores de una revisión deben llevar a cabo estrategias para detectar estos posibles sesgos, así como investigar su posible relación con el efecto de la intervención y la precisión de la estimación.

Diferencias principales entre QUOROM y PRISMA

La nueva lista de comprobación PRISMA difiere en varios aspectos con respecto a QUOROM. En primer lugar, PRISMA presenta un total de 27 ítems (tabla 1) frente a los 18 ítems de QUOROM.

Algunos ítems se han incorporado de nuevo a la lista, mientras que otros resultan del desdoblamiento de ítems ya presentes en QUOROM. Además, algunos ítems están interrelacionados para mejorar la consistencia en el informe de la revisión sistemática por parte de los autores.

Los cambios más relevantes introducidos en la lista de comprobación de PRISMA se presentan en la tabla 2.

El diagrama de flujo también se ha modificado, y este nuevo diagrama de PRISMA es más detallado e informativo (fig. 1). El diagrama de QUOROM partía de los ensayos clínicos aleatorizados potencialmente relevantes sometidos a cribado y terminaba con los ensayos que proporcionan información útil para cada resultado en el metaanálisis, mientras que el de PRISMA parte desde el inicio mismo del proceso (los registros o las citas identificados en las búsquedas realizadas en cada una de las diferentes bases de datos u otras fuentes utilizadas), continúa por el número total de registros o citas únicas una vez eliminados los duplicados y termina con los estudios individuales incluidos

Tabla 1

Lista de comprobación de los ítems para incluir en la publicación de una revisión sistemática (con o sin metaanálisis). La declaración PRISMA

Sección/tema	Número	Ítem
<i>Título</i> Título	1	Identificar la publicación como revisión sistemática, metaanálisis o ambos
<i>Resumen</i> Resumen estructurado	2	Facilitar un resumen estructurado que incluya, según corresponda: antecedentes; objetivos; fuente de los datos; criterios de elegibilidad de los estudios, participantes e intervenciones; evaluación de los estudios y métodos de síntesis; resultados; limitaciones; conclusiones e implicaciones de los hallazgos principales; número de registro de la revisión sistemática
<i>Introducción</i> Justificación	3	Describir la justificación de la revisión en el contexto de lo que ya se conoce sobre el tema
Objetivos	4	Plantear de forma explícita las preguntas que se desea contestar en relación con los participantes, las intervenciones, las comparaciones, los resultados y el diseño de los estudios (PICOS)*
<i>Métodos</i> Protocolo y registro	5	Indicar si existe un protocolo de revisión al que se pueda acceder (por ej., dirección web) y, si está disponible, la información sobre el registro, incluyendo su número de registro
Criterios de elegibilidad	6	Especificar las características de los estudios (por ej., PICOS, duración del seguimiento) y de las características (por ej., años abarcados, idiomas o estatus de publicación) utilizadas como criterios de elegibilidad y su justificación
Fuentes de información	7	Describir todas las fuentes de información (por ej., bases de datos y períodos de búsqueda, contacto con los autores para identificar estudios adicionales, etc.) en la búsqueda y la fecha de la última búsqueda realizada
Búsqueda	8	Presentar la estrategia completa de búsqueda electrónica en, al menos, una base de datos, incluyendo los límites utilizados, de tal forma que pueda ser reproducible
Selección de los estudios	9	Especificar el proceso de selección de los estudios (por ej., el cribado y la elegibilidad incluidos en la revisión sistemática y, cuando sea pertinente, incluidos en el metaanálisis)
Proceso de extracción de datos	10	Describir los métodos para la extracción de datos de las publicaciones (por ej., formularios pilotado, por duplicado y de forma independiente) y cualquier proceso para obtener y confirmar datos por parte de los investigadores
Lista de datos	11	Listar y definir todas las variables para las que se buscaron datos (por ej., PICOS, fuente de financiación) y cualquier asunción y simplificación que se hayan hecho
Riesgo de sesgo en los estudios individuales	12	Describir los métodos utilizados para evaluar el riesgo de sesgo en los estudios individuales (especificar si se realizó al nivel de los estudios o de los resultados) y cómo esta información se ha utilizado en la síntesis de datos
Medidas de resumen	13	Especificar las principales medidas de resumen (por ej., razón de riesgos o diferencia de medias)
Síntesis de resultados	14	Describir los métodos para manejar los datos y combinar resultados de los estudios, cuando esto es posible, incluyendo medidas de consistencia (por ej., ítem 2) para cada metaanálisis
Riesgo de sesgo entre los estudios	15	Especificar cualquier evaluación del riesgo de sesgo que pueda afectar la evidencia acumulativa (por ej., sesgo de publicación o comunicación selectiva)
Análisis adicionales	16	Describir los métodos adicionales de análisis (por ej., análisis de sensibilidad o de subgrupos, metarregresión), en el caso de que se hiciera, indicar cuáles fueron preespecificados
<i>Resultados</i> Selección de estudios	17	Facilitar el número de estudios cribados, evaluados para su elegibilidad e incluidos en la revisión, y detallar las razones para su exclusión en cada etapa, idealmente mediante un diagrama de flujo
Características de los estudios	18	Para cada estudio presentar las características para las que se extrajeron los datos (por ej., tamaño, PICOS y duración del seguimiento) y proporcionar las citas bibliográficas

Tabla 1 (continuación)

Sección/tema	Número	Ítem
Riesgo de sesgo en los estudios	19	Presentar datos sobre el riesgo de sesgo en cada estudio y, si está disponible, cualquier evaluación del sesgo en los resultados (ver ítem 12)
Resultados de los estudios individuales	20	Para cada resultado considerado en cada estudio (beneficios o daños), presentar: a) el dato resumen para cada grupo de intervención y b) la estimación del efecto con su intervalo de confianza, idealmente de forma gráfica mediante un diagrama de bosque (<i>forest plot</i>)
Síntesis de los resultados	21	Presentar los resultados de todos los metaanálisis realizados, incluyendo los intervalos de confianza y las medidas de consistencia
Riesgo de sesgo entre los estudios	22	Presentar los resultados de cualquier evaluación del riesgo de sesgo entre los estudios (ver ítem 15)
Análisis adicionales	23	Facilitar los resultados de cualquier análisis adicional, en el caso de que se hayan realizado (por ej., análisis de sensibilidad o de subgrupos, metarregresión [ver ítem 16])
<i>Discusión</i>		
Resumen de la evidencia	24	Resumir los hallazgos principales, incluyendo la fortaleza de las evidencias para cada resultado principal; considerar su relevancia para grupos clave (por ej., proveedores de cuidados, usuarios y decisores en salud)
Limitaciones	25	Discutir las limitaciones de los estudios y de los resultados (por ej., riesgo de sesgo) y de la revisión (por ej., obtención incompleta de los estudios identificados o comunicación selectiva)
Conclusiones	26	Proporcionar una interpretación general de los resultados en el contexto de otras evidencias, así como las implicaciones para la futura investigación
<i>Financiación</i>		
Financiación	27	Describir las fuentes de financiación de la revisión sistemática y otro tipo de apoyos (por ej., aporte de los datos), así como el rol de los financiadores en la revisión sistemática

* PICOS: se trata de un acrónimo formado por: P: participants; I: interventions; C: comparisons; O: outcomes; S: study design.

Tabla 2

Cambios más relevantes introducidos en la lista de comprobación de PRISMA

Sección/tema	Ítem	Comentario
Título		PRISMA solicita la identificación de la publicación como revisión sistemática, metaanálisis o ambos, mientras que QUOROM se refería exclusivamente a metaanálisis de ensayos clínicos
Resumen		Tanto QUOROM como PRISMA solicitan a los autores un resumen estructurado, pero PRISMA no especifica su formato, aunque sí hace recomendaciones
Introducción	Objetivo	Este nuevo ítem (4) se focaliza en la pregunta explícita que aborda la revisión utilizando el formato PICO (descripción de los participantes, las intervenciones, las comparaciones y las medidas de resultado de la revisión sistemática), así como el tipo de estudio (diseño); este ítem está interrelacionado con los ítems 6, 11 y 18
Métodos	Protocolo	Este nuevo ítem (5) solicita al autor que explique si un protocolo precedió la revisión y, en tal caso, cómo puede accederse a éste
Métodos	Búsqueda	PRISMA desdobra el ítem sobre la «búsqueda» de QUOROM en 2 ítems: a) fuentes de información (ítem 7) y b) búsqueda (ítem 8). Aunque la estrategia de búsqueda se reporta tanto en QUOROM como en PRISMA, éste solicita a los autores que proporcionen la descripción completa de, al menos, una estrategia de búsqueda electrónica (ítem 8), sin ésta no es posible reproducir la búsqueda
Métodos	Evaluación del riesgo de sesgo en los estudios incluidos	En QUOROM este ítem aparecía como «evaluación de la calidad». Ahora, este ítem (12) se focaliza en la evaluación del riesgo de sesgo dentro de cada estudio incluido en la revisión. Además, este ítem está interrelacionado con otro nuevo ítem incorporado en PRISMA: la comunicación de esta información en los resultados (ítem 19). También se introduce el nuevo concepto de evaluación del sesgo al nivel de los resultados

PRISMA: Preferred Reporting Items for Systematic reviews and Meta-Analyses.

en la síntesis cualitativa (revisión sistemática) y cuantitativa (metaanálisis). PRISMA también establece la diferencia en cada etapa del proceso entre los registros o las referencias bibliográficas (resultado de aplicar las estrategias de búsqueda electrónica en las bases bibliográficas), los artículos a texto completo (artículos que deben obtenerse a texto completo para decidir con seguridad acerca de su elegibilidad o no) y los estudios individuales (estudios que cumplen los criterios de elegibilidad de la revisión y que pueden corresponderse con una o más publicaciones o artículos).

En cuanto a las expectativas para el futuro, PRISMA se ha concebido como una herramienta para contribuir a mejorar la claridad y la transparencia en la publicación de revisiones sistemáticas. Por el contrario, PRISMA no se ha formulado como

un instrumento para valorar la calidad de las revisiones y no debería utilizarse como tal.

Como los autores señalan, seguir las recomendaciones de la lista de comprobación PRISMA puede suponer aumentar la extensión de la publicación de una revisión sistemática, lo que puede colisionar con las normas editoriales de algunas revistas. No obstante, las ventajas de proporcionar a los lectores información completa, clara y transparente superan los inconvenientes de tener que leer un texto algo más extenso.

Si las revistas biomédicas van a adoptar la declaración PRISMA, tal como ya lo fueron otras propuestas como CONSORT, debería existir evidencia que demuestre que PRISMA realmente mejora la claridad y la transparencia de las revisiones sistemáticas publicadas. Los autores de PRISMA tienen la intención de evaluar

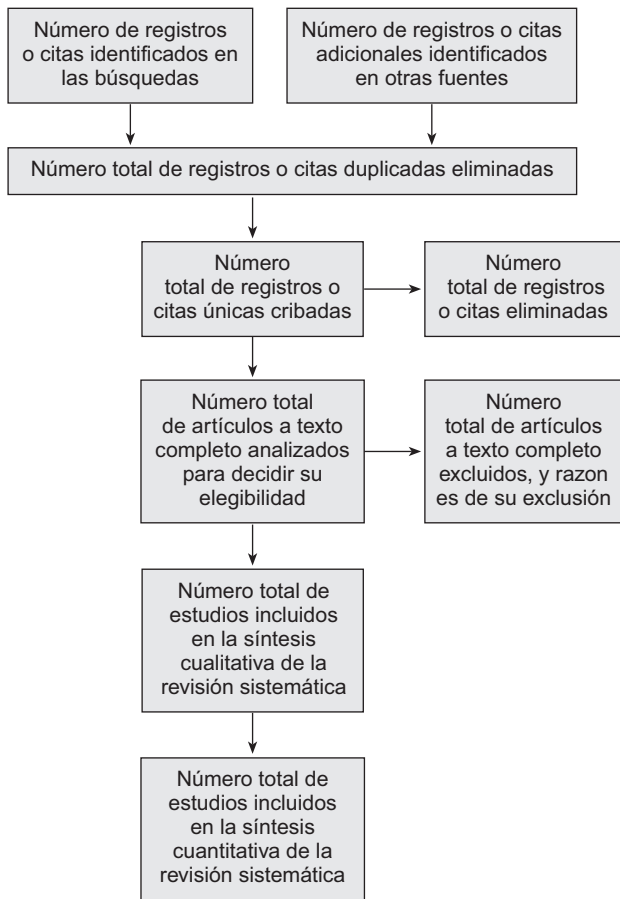


Figura 1. Diagrama de flujo de la información a través de las diferentes fases de una revisión sistemática.

los beneficios (calidad mejorada de la publicación) y los posibles efectos adversos (aumento de la extensión del texto) de PRISMA y animan a otros a hacer lo mismo.

Una de las limitaciones apuntadas por los autores de PRISMA es que no se realizó una revisión sistemática previa para la confección de la lista de comprobación. No obstante, PRISMA se desarrolló sobre la base de las evidencias disponibles siempre que fue posible. Solo se incluyeron en el listado ítems para los que hubiera evidencia de una relación entre su omisión y el riesgo de sesgo, o bien que existiera amplio consenso acerca de su necesidad para evaluar la fiabilidad de una revisión.

Con el fin de mantener PRISMA actualizado y basado en la evidencia, los autores realizarán una revisión periódica de la literatura médica metodológica. Por ejemplo, no existen evidencias publicadas para apoyar algunos ítems, como por ejemplo, si el entrenamiento previo contribuye a mejorar la precisión y la fiabilidad en la extracción de los datos. Por esto, se espera que PRISMA actúe de catalizador de futuros estudios que generen la evidencia que pueda incorporarse en futuras actualizaciones del listado de comprobación.

Los autores confían en que PRISMA tendrá una revisión y una actualización más frecuentes, así como también una mejor implementación que la que tuvo en su momento QUOROM. Este optimismo se basa en el número creciente de revisiones sistemáticas publicadas y su utilización, también creciente, por parte de los proveedores, los decisores y los gestores en salud para fundamentar las decisiones clínicas, sanitarias y de investigación. También basan su optimismo en los beneficios esperados del desarrollo de la Red EQUATOR (Enhancing the QUALity and Transparency Of health Research), cuyo objetivo es ayudar a individuos y a grupos interesados en el desarrollo de futuras directrices (o su traducción) dirigidas a mejorar la calidad de las publicaciones en ciencias de la salud^{10,11}. La red dispone de una web donde se facilita el acceso a los recursos (<http://www.equator-network.org/>).

El éxito de PRISMA está por verse, y dependerá de la medida en que las revistas biomédicas y los grupos editoriales la adopten y la incluyan en las «Instrucciones para autores», donde se establezca un *link* con la web de PRISMA, y realicen acciones editoriales para llamar la atención de los lectores y los potenciales autores de revisiones acerca de su importancia.

Conflicto de intereses

Los autores declaran no tener ningún conflicto de intereses.

Bibliografía

- Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF, et al. Improving the quality of reports of meta-analyses of randomised controlled trials: The QUOROM statement. *Lancet*. 1999;354:1896-900.
- Urrutia G, Tort S, Bonfill X. Metaanálisis (QUOROM). *Med Clin (Barc)*. 2005;125:32-7.
- Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and reporting characteristics of systematic reviews. *PLoS Med*. 2007;4:e78, doi:10.1371/journal.pmed.0040078.
- Delaney A, Bagshaw SM, Ferland A, Manns B, Laupland KB, Doig CJ. A systematic evaluation of the quality of meta-analyses in the critical care literature. *Crit Care*. 2005;9:R575e82.
- Site web de PRISMA. [consultado 23/02/2010] Disponible en: <http://www.prisma-statement.org>.
- Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration. *BMJ*. 2009;339:b2700, doi:10.1136/bmj.b2700.
- Green S, Higgins JPT, Alderson P, Clarke M, Mulrow CD. Chapter 1: What is a systematic review?. [actualizado 2/2008]. En: Higgins JPT, Green S, editores. *Cochrane handbook for systematic reviews of interventions version 5.0.0*. The Cochrane Collaboration; 2008 [consultado 26/11/2009]. Disponible en: <http://www.cochrane-handbook.org/>.
- Higgins JPT, Green S, editores. *Cochrane handbook for systematic reviews of interventions version 5.0.0*. The Cochrane Collaboration; 2008 [consultado 26/11/2009]. Disponible en: <http://www.cochrane-handbook.org/>.
- Higgins JPT, Altman DG. Chapter 8: Assessing risk of bias in included studies. [actualizado 2/2008]. En: Higgins JPT, Green S, editores. *Cochrane handbook for systematic reviews of interventions version 5.0.0*. The Cochrane Collaboration; 2008 [consultado 26/11/2009]. Disponible en: <http://www.cochrane-handbook.org/>.
- Altman DG, Simera I, Hoey J, Moher D, Schulz K. EQUATOR: Reporting guidelines for health research. *Lancet*. 2008;371:1149e50.
- Simera I, Altman DG, Moher D, Schulz KF, Hoey J. Guidelines for reporting health research: The EQUATOR network's survey of guideline authors. *PLoS Med*. 2008;5:e139, doi:10.1371/journal.pmed.0050139.

Clinical Research

Is the Spine Field a Mine Field?

Eugene J. Carragee, MD,* Richard A. Deyo, MD, MPH,† Francisco M. Kovacs, MD, PhD,‡§
 Wilco C. Peul, MD, PhD,¶|| Jon D. Lurie, MD, MS,|| Gerard Urrútia, MD,**
 Terry P. Corbin, BS,†††† and Mark L. Schoene, BS§§

Spinal surgery for infection, tumor, and major traumatic disruption may avert catastrophic outcomes such as paralysis, major deformity, or death. Recent research on decompression and fusion surgery for degenerative spinal disorders has shown the clear and relevant effectiveness of many procedures. When clinically indicated, lumbar discectomy for disc herniation, decompression for lumbar stenosis, and decompression and fusion surgery for degenerative spondylolisthesis, all offer clinical benefit in the face of serious back and radicular pain.^{1–5}

The scientific basis for these treatments began with one of the earliest randomized trials of a surgical intervention,⁶ establishing the spine field as a leader in innovative research methods in surgery. The results of that early study have been largely confirmed in more recent randomized trials.^{7–9}

Technical advances have made many spine operations and other interventions safer and more predictable in outcome. These advances have greatly improved critical aspects of care, including antibiotic and thromboembolic disease prophylaxis, intraoperative imaging, lighting and magnification, and analgesic strategies for pain. Critical spinal devices have improved in biocompatibility, durability, and simplicity of implantation. These advances are generally the product of cooperative research and development between the innovative clinician, clinical researcher, and the biomedical industry. The time when a clinician with a good idea could bring a new device from workbench to practice without enormous commercial

and capital support is long gone. Physicians, patients, and industry share many common practical goals.

However, not all new treatment strategies prove to be better or even safe. Early trials with several techniques and devices reveal a disturbing history (Table 1). Nonetheless, new and potentially helpful devices are being proffered to the spine care market at an accelerating pace. Each year, the number of new products available for consideration outstrips the ability of any individual to adequately assess their safety, effectiveness, or even reasonableness. In practice, the spine field has relied on regulatory agencies, university, and unaffiliated research programs and unbiased scientific reporting in the literature to help sort the wheat from the chaff. There are signs that this method of assessing new technology has been broken for some time and, as a consequence, clinicians concerned with spinal disorders are facing hard times ahead unless a better strategy is found.

Over the past decade, the rate of some spinal procedures being performed in North America and Europe has greatly increased.¹⁰ For instance, a rapid increase in fusion rates followed the approval of new surgical implants, but was not associated with evidence of improved effectiveness or clarified indications.¹¹

Many interventions are performed for axial back pain associated with common degenerative conditions, sometimes with weak or absent evidence of efficacy: so-called “dynamic stabilization” with various implants, distraction with interspinous devices, corticosteroid injection into the facet joints, prolotherapy, intradiscal steroid injection, botulinum toxin injection, intradiscal electrothermal therapy, radiofrequency facet denervation, radiofrequency sacroiliac denervation, and intrathecal therapy with opioids, to name a few. Though thousands of patients have been treated with these methods, there is no evidence of high-grade efficacy for even one of them.

At the same time, highly publicized news stories of surgeons receiving “millions in consulting fees, royalties and research grants” have appeared, in the context of “sham consulting agreements, sham royalty agreements, and lavish trips to desirable locations.”¹² A large device manufacturer has paid a fine to the government to settle such allegations.¹³ As a result, the spine field has been subject to ugly characterization in the press, highly skeptical governmental investigations, and frank accusations of corruption. *Consumer Reports* recently listed spinal surgery as number 1 on its list of overused tests and treatments.¹⁴ This is clearly not where this field wants to be.

From the *Department of Orthopaedic Surgery, Stanford University School of Medicine, Palo Alto, CA; †Department of Family Medicine, Oregon Health and Science University, Portland, OR; ‡Research Department, Kovacs Foundation, Palma de Mallorca, Spain; §Spanish Back Pain Research Network; ¶Department of Neurosurgery, Leiden University Medical Center, Leiden, The Netherlands; ||Department of Medicine, Dartmouth Medical School, Hanover, NH; **Iberoamerican Cochrane Center, Servei d’Epidemiologia i Salut Pública, Hospital de la Santa Creu i Sant Pau, Universitat Autònoma de Barcelona, Barcelona, Spain; ††University of Minnesota, Clinical Outcomes Research Center, Minneapolis, MN; †††Corbin and Company; and §§The Back Letter, Newbury, MA.

Acknowledgment date: September 25, 2008. Acceptance date: December 10, 2008.

The manuscript submitted does not contain information about medical device(s)/drug(s).

No funds were received in support of this work. No benefits in any form have been or will be received from a commercial party related directly or indirectly to the subject of this manuscript.

Address correspondence and reprint requests to Eugene J. Carragee, MD, FACS, Department of Orthopaedic Surgery, Stanford University School of Medicine; Orthopaedic Spine Center and Spinal Surgery Service, Stanford University Hospital and Clinics, Stanford, CA; E-mail: ecarragee@mac.com

Table 1. Some Technologies That Have Been Widely Disseminated, but With Conflicting Clinical Trials or Seemingly Modest Benefits

Technology/Device	Current Status of Evidence and Practice
Intradiscal electrothermal therapy (IDET)	Multiple favorable cases series. Approved on basis of unpublished case series of 25 patients. An independent systematic review including all the available RCTs concluded that “the evidence does not support the efficacy or effectiveness of IDET for the treatment of discogenic low back pain.” ⁵⁸ Approximately 7000–10,000 IDET procedures are performed annually in the United States, where more than 75,000 procedures have been performed since 1998. (http://www.idetprocedure.com/2000_professional/2100_about.html , visited on Nov 12, 2007)
Automated percutaneous discectomy	Multiple favorable case series. However, 1 RCT showed results were inferior to chemonucleolysis. And multiple RCTs suggest chemonucleolysis is inferior to standard discectomy. ⁵⁹ A Cochrane Collaboration review found no evidence to support the use of automated percutaneous discectomy in routine clinical practice. ⁶⁰
Laser discectomy	Multiple favorable case series. Yet 3 RCTs found no conclusive evidence of its efficacy, according to a Cochrane Collaboration review. ⁶¹ No longer in wide use.
Adcon-L, antiadhesion gel for spine surgery	Approved by FDA in 1997, recalled in United States in 2001. In 2002, company admitted to providing false or incomplete information to FDA. FDA raised concerns over rescoring of MRI scans used to measure scar tissue, and “erasures and writeovers” of data. Company also received reports of adverse reactions, which it failed to report to FDA as required. The manufacturer was ultimately forced into bankruptcy. ⁶¹
Stand-alone interbody fusion cages	Approved based solely on case series. Nonindustry-sponsored studies showed less favorable results than industry-sponsored case series. ⁶² Never proven superior to other treatments in RCTs. Described by president of 1 spine society as prompting a “fusion cage explosion,” but “4 yr later, the efficacy of these stand-alone devices is very questionable.” ⁶³
Vax-D, DRX9000 and similar spinal traction/decompression devices	Widely publicized and disseminated. Despite the publication of 2 RCTs on spinal decompression therapy, a 2007 systematic review published by the Agency for Healthcare Quality and Research could find no conclusive evidence of its effectiveness. “Currently available evidence is too limited in quality and quantity to allow for the formulation of evidence-based conclusions regarding the efficacy of decompression therapy as a therapy for chronic back pain . . .” ⁶⁴

Sources of the Problem

Limitations of Food and Drug Administration (FDA) Regulation.

FDA regulation of medical devices is limited, and has only existed since 1976. FDA officials have acknowledged that “New devices are less likely than drugs to have their safety established clinically before they are marketed.”¹⁵ Considering the track record of Vioxx and recent prosecution over the addictive potential of OxyContin, this is alarming. Similarly, the regulatory process does not assure that new devices are better than old ones, that ineffective devices are removed from the market, that new devices are systemati-

cally monitored after approval, or that devices offer good value for money (Table 2). Furthermore, internal and external pressures for rapid approval of new products have been described, and conflicts of interest among members of FDA advisory committees are well documented.¹⁶ These limitations and pressures help to explain why ineffective devices may sometimes be marketed, and why insurers increasingly demand more extensive evidence of efficacy and safety before coverage decisions.¹⁷ In Europe, despite differences in the organization of health care and coverage, the situation is very similar.

Table 2. Myths and Realities About Medical Device Evaluation and Regulation

Myth	Reality
New device must be better than an old one to be approved	No comparison with competing devices is required for approval; a new device may be equally effective, more effective, or less effective than alternatives.
Studies done for approval are likely to uncover all important side effects	For new drugs, 50% have important side effects discovered only after approval and marketing. No data available for devices, but gaps in the approval process suggest likelihood of a problem of at least similar magnitude.
Device approval is similar to drugs	Most devices (over 98%) approved on a claim of “substantial equivalence” to something already on the market. This may seem paradoxical, as marketing usually emphasizes newness and innovation. Evaluation focuses on safety and engineering performance rather than clinical efficacy. Unlike drugs, no randomized trials required except for devices judged to be the most innovative or risky.
Devices are approved only if they offer good value for money	FDA is prohibited from considering cost or cost-effectiveness in the approval process.
Off-label uses of devices are well-evaluated	Off-label use of many devices is common and legal, but efficacy and safety are often poorly established because there is no incentive to study after a device is approved for at least 1 indication.
Ineffective devices are recalled	Devices are recalled only for safety problems and not for accumulating evidence of poor efficacy.
Surgical procedures are regulated by FDA	No procedures are regulated by the FDA; only devices are regulated.

Table 3. Strategies for Clinicians and Researchers in Improving Device Evaluations

Stakeholder, Strategy	Details, Rationale
Clinicians	
Become familiar with principles of evidence-based medicine	Understand hierarchy of rigor of efficacy data; critically evaluate new evidence; recognize that newer is not necessarily better.
Promote shared decision-making with patients	Patients often do not realize most spine procedures are elective; may not be familiar with alternatives; may not understand the magnitude of expected benefits from each choice; better information reduces medicolegal risks.
Avoid consulting arrangements that are disproportionate to effort	Financial interest may influence choice of implants or procedure, rather than patient interests alone; adverse publicity is eroding public confidence and trust.
Encourage well informed patient participation in rigorous research	High levels of participation are needed for successful clinical trials. For most new products, there is genuine uncertainty about comparative effectiveness, making clinical studies both necessary and ethical.
Avoid use in applications that are different from those used in the assessment process (<i>e.g.</i> , off-label use of devices, differences in training standards, <i>etc.</i>)	Especially for a nonpharmacologic technology, evidence on safety and effectiveness is not valid if application conditions (including indication criteria and training standards) are different from those under which it was assessed. Unless rigorous clinical trials have been conducted, use in off-label applications amounts to experimentation without consent.
Researchers, authors	
Avoid ghost authorship	For industry-sponsored studies, ghostwriting by the manufacturer or a public relations firm is common. It is likely to present and emphasize only the most favorable results, and offer the most favorable interpretation. Publication is a key aspect of many marketing plans. "Big names" are sometimes featured as authors to enhance credibility.
Avoid ghost management of research	Industry-sponsored studies are aimed at device approval, not at the most clinically important questions (<i>e.g.</i> , is this device better or worse than the optimal alternative?).
Ensure that researchers/authors make all publication decisions	Unless clear in contracts, sponsors can choose not to publish unfavorable results, often citing "proprietary" nature of data. Suppression and delay of unfavorable results are extensively documented in the drug industry.
Retain control of all data; Perform all data analyses according to the pre-hoc analysis plan included in the research protocol	Unless investigators control and analyze data, arbitrary decisions can be made that produce the most favorable "spin" on results. JAMA insists that at least 1 named author who is independent of any commercial funding source has full access to all data and takes responsibility for integrity of the data and accuracy of the analysis. For industry-sponsored studies, JAMA requires data analysis by an independent statistician at an academic institution or government research institute that is independent of the commercial sponsor, with report of any compensation received for the analysis. Investigators should make data available to the research community (<i>e.g.</i> , on a website), to make it possible to confirm results and perform further analyses. ^{65,66}

Incongruence Between Industry Goals and Clinical Goals.

Trials intended to win FDA approval of new products are designed to provide some evidence of safety and effectiveness. However, clinicians seek novel products that are more effective than existing treatments. Whether a new device truly has superior properties or cost-effectiveness in comparison with the best alternative treatment is rarely investigated. Clinical adoption of new spine devices—both in the United States and elsewhere—is often faster than the publication of high-quality evidence, and especially of randomized controlled trials.

In some cases, industry-sponsored trials fail to compare the novel device with a state-of-the-art alternative. The FDA study of the CHARITÉ artificial disc for low back pain¹⁸ compared it with a surgical "standard" (fusion with stand-alone cages), which many would not consider state-of-the-art surgical management of discogenic pain.¹⁹ Stand-alone fusion cages were themselves never subjected to randomized trials before approval, leaving an obvious gap in the chain of evidence.

Some industry-sponsored studies, by their design, fail to address major clinical issues. For example, the rate of adjacent level disease after cervical spine fusion is relatively high (up to 25% after 10 years).²⁰ Manufacturers

argue that cervical disc prostheses can prevent adjacent level disease in the years after surgery. To address this clinical concern, randomized trials should aim at a comparison of the cervical disc prosthesis with state-of-the-art fusion or discectomy alone, with a large enough sample size (and adequate statistical power) to answer the research question. Such trials have not been performed, yet potential dissemination of these devices worldwide is enormous.^{21,22}

The goals of industry are somewhat different than those of academic medicine or practicing clinicians. Concerns about these differences are growing as, in some areas, for-profit contract research organizations have taken over much of academia's traditional role in clinical research.²³

Industry funding of research may lead to publication bias. Among studies published in *Spine*, those with industry funding were more likely to show positive results than independent studies.²⁴ Clinicians should be aware of this problem, and independent academic spine researchers should be intimately involved in the design, conduct, and analysis of randomized controlled trials. Cooperation with industry is necessary, but with rigorous safeguards, transparency of design and analysis, and with practical clinical goals in mind (Table 3).

Premature Distribution and Marketing. Many controversial devices have been marketed in recent years. In some cases, they experienced rapid dissemination, followed by disillusionment and a decline in use. Often, when randomized trials have been belatedly performed, the results have been conflicting or suggest only meager benefits (Table 1). In the short run, such devices may be profitable, perhaps explaining why they are marketed in the absence of rigorous evidence of their benefit. But in the long run, these patterns do not serve patients, physicians, or industry well.

Withholding Unfavorable Results or Access to Data. Presentations and abstracts at scientific meetings may influence practice, but more than half are never published in full reports for scientific scrutiny.²⁵ Negative trials may not be published, with corporate sponsors arguing the data are proprietary. In the case of some drugs, such as Vioxx and OxyContin, it seems that data on significant side effects were known to the manufacturers well before they were made widely available. Contracts for industry-sponsored research often include clauses leaving decisions about publication of results up to the sponsor.²⁶ Sometimes, data from clinical studies are presented in piecemeal fashion, making it difficult to assemble a complete picture of all the results. Recent publications of selected centers of multicenter FDA artificial disc trials demonstrate this problem.^{27–34}

For example, Sasso *et al* published results of the trial comparing the Bryan disc to ACDF from only 3 of the 31 centers participating in the study.³⁵ The authors did not report the full range of surveillance for complications. In fact, they reported only 6 complications, adverse events, or reoperations among 99 patients. Remarkably, no subject in either group was noted to have dysphagia, hoarseness, or perioperative neurologic complications.

Yet a subsequent study of all 463 patients at the 31 centers documented more than 200 complications and adverse events.³⁶ In that study, 42 patients reported dysphagia or hoarseness, and another 15 reported perioperative neurologic problems.

As these studies demonstrate, piecemeal reporting of study results can be misleading. It also raises statistical concerns. For example, single-center piecemeal reporting on groups of 20 to 30 subjects may not be adequately powered to detect even 10% to 15% rates of catastrophic complications or failure. Yet, claims of safety are frequently made in peer-reviewed journals based on small numbers of subjects.^{37,38}

Even with a relatively large FDA trial of 200 subjects, such as the CHARITÉ RCT, patients receiving such a device cannot assume the risk of catastrophic failure or complication is less than 1% to 2% with 95% confidence interval. Only with wider application can the true risks be estimated, and at that point systematic surveillance is often lacking.

Very often safety concerns are addressed only marginally in the approval process, and the level of risk is often

not appreciated by clinicians applying the approved device in practice. For instance, osteoporosis was a contraindication to the implantation of lumbar artificial discs in the FDA trials of both the CHARITÉ disc and the Prodisc.^{39,40} Yet, the chances are high that 50-year-old woman receiving these implants will eventually develop levels of osteoporosis, which are assumed to be incompatible with these prostheses.

In some cases, corporate sponsors control all the data and analyses, which are not made available to the investigator. One researcher studying osteoporosis at a UK university had the disconcerting experience of seeing his name on abstracts submitted to a scientific meeting without ever seeing the study data. When partial data were made available, he reported that there were extensive missing values.^{41,42}

Professional Societies and Conflicts of Interest. A brief look at the exhibit hall or at the programs of professional meetings illuminates the importance of industry support for many organizations. This can create conflicts of interest for professional societies, analogous to conflicts for individual clinicians. As a neurosurgeon editor remarked: “Since the device companies provide generous support of our national meetings, are our national leaders going to take a position that may offend those companies? No, not a chance.”⁴³ Because no one wants to bite the hand that feeds him (or her), professional criticism of dubious research or marketing practices is often muted.

Growing Dissatisfaction With the Spine Research System

There is growing dissatisfaction with the current research system—and for good reason. Haldeman and Dagenais recently pointed out that clinicians and patients face an increasingly confusing array of treatments for chronic back pain. The inventory of treatment options includes more than 200 different medications, therapies, injections, products, and procedures, and is growing continually. Yet many of these treatment approaches lack scientific validation because of basic flaws in the research process.⁴⁴

Major payers are voicing increasing skepticism about the quality and fairness of scientific research on spinal treatments and technologies. The Centers for Medicare and Medicaid Services limited coverage for lumbar disc replacement because of problems in the design and conduct of commercially sponsored clinical trials.⁴⁵ In a more recent proposed noncoverage determination on thermal intradiscal procedures, Centers for Medicare and Medicaid Services expressed the concern that financial conflicts of interest may be skewing the results of both clinical and basic research in favor of treatments supported by industry.⁴⁶

A recent commentary in JAMA labeled the current system of industry-sponsored clinical research as “broken”—and suggested that clinicians can no longer rely

Table 4. Regulatory Strategies for Improving Device Evaluation

Strategy	Rationale
Require randomized trials of new surgical implants or devices that destroy or alter tissue	Case series provide no comparison with best alternative standard care. Nonrandomized trials risk having important but unmeasured differences between study groups. Nonrandomized designs are useful for establishing safety, but are inadequate for evaluating clinical efficacy (improvement in pain and function).
Require comparative effectiveness studies of competing devices	If we believe every device is equivalent to all competitors, there is no need for changing designs or for innovations. If we think there is an advantage of a particular design, it should be rigorously demonstrated so that all patients and surgeons can benefit.
Systematic postmarketing surveillance	Public or governmental agencies, such as the FDA in the United States, have few resources for post-marketing surveillance. As a result, systematic postmarketing surveillance is rarely required. In practice, agencies such as the FDA have little authority to enforce their requests for postmarketing studies.
Independent device safety board	Plane crashes are investigated by someone other than those who approve and operate the planes and the air traffic control system. Device safety evaluation demands a similar level of independence from those who approve, make, or use the devices.
Consider conditional device approvals	Phased introduction of new devices would allow an assessment of safety and effectiveness on a larger scale than typically small clinical studies—before more widespread use.
Reduce conflicts of interest on advisory boards	Individuals with financial interests in the success of a device should not have a role in its approval.

on the medical literature for valid and reliable information.⁴⁷

Near-Term Consequences

It is an understatement to point out that resources for health care are limited. On its current trajectory, Medicare is forecast to become bankrupt without new restrictions on benefits or new taxes. Auto manufacturers are reducing health care benefits to avoid bankruptcy. Some 46 million Americans have no health insurance, because it has become unaffordable. The Director of the Congressional Budget Office, writing in the *New England Journal of Medicine*, links the future of the entire US economy to controlling the growth of health care costs. He notes that the bulk of spending growth is not from increasing disease prevalence or aging, “but from the development and diffusion of new medical technologies and therapies.” He further notes, “substantial evidence exists that more expensive care doesn’t always mean higher quality care.”⁴⁸

In this context, every dollar spent on ineffective or marginal treatments is a dollar less that is available for highly effective health care. Ethics, as well as economics, argue against providing marginal treatments, when greater health benefits can be gained by focusing on highly effective treatments. Unless physicians take a role in objectively sorting treatments with good value from ineffective care, there is a risk of indiscriminate cuts in services. Administrators and politicians do not have the scientific skills to distinguish the good from the marginal. If, for example, overuse of spinal fusion surgery for dubious indications led to blanket restrictions that applied to use in fractures, tumors, infections, or deformities, our ability to deliver some of the most effective treatments in spinal disorders may be impeded.

Similarly, if conflicts of interest in research and marketing lead to draconian restrictions on physician-industry cooperation, true innovative progress may be stifled along with frivolous novelty. This cooperation is

essential, but transparency about financial arrangements and data management is needed to ensure the integrity of clinical research.

Recommendations to Avert Disaster

The authors of this commentary believe changes are needed on the part of all stakeholders: clinicians, researchers, regulators, insurers, and device manufacturers (Tables 3–5). These proposed changes follow directly from our observations about the sources of the problem, and many of these suggestions have been advocated elsewhere. The changes have to do with transparency of financial arrangements, publication policies, research design and data management, device approval and post-marketing surveillance, and even support for research. The Tables 3 to 5 present our suggestions and their rationales.

These suggestions build on prior proposals in the medical literature. The American Academy of Orthopedic Surgeons recently proposed new standards to govern the disclosure and management of financial conflicts of interest between orthopedic surgeons and industry—and to prevent those conflicts from having an adverse effect on scientific research, physician education, and patient care.⁴⁹ They provide a good starting point for further debate on these issues.

Though much remains to be done, some modest changes have begun in the regulatory process.⁵⁰ New initiatives will open up a vast quantity of research on new treatments and technologies to scientific and public scrutiny. The 2007 Food and Drug Administration Amendment Act requires open access to both efficacy and safety data from trials of FDA-approved drugs, biologics, and devices (except Phase I trials), generally within a year of the trials’ completion—whether the trials are published or not.⁵¹ Other groups have recommended that the summary results of all clinical trials on new treatments and

Table 5. Strategies for the Manufacturing and Insurance Industries to Improve Device Evaluation

Stakeholder, Strategy	Rationale
Device industry	
Encourage independent data collection and analysis in clinical research studies, independent publication decisions	Same as in Table 3.
Preregister all clinical studies	Registering all study designs and outcome measures in advance helps to assure that all results are made available, not just favorable results. ^{65,66} Many journals will now only publish randomized trials that have been preregistered, <i>e.g.</i> , on an NIH website.
Publish (<i>e.g.</i> , on a website) all payments to physicians, scientists, and institutions who are not company employees	Transparency is the best way to deal with conflicts of interest, which are inevitable.
Insurance industry	
Provide financial support for independent trials of selected new technologies	Research studies are expensive, but not nearly as expensive as paying for years of ineffective treatments and related complications.
Consider evidence-based coverage decisions	This can prevent 2 unfortunate situations: refusing coverage simply because a treatment is expensive; or paying for treatments that do not work. Because of gaps and conflicts in FDA approval process, approval alone is not an adequate justification for coverage. Evidence-based coverage requires independent and transparent assessment of new technology by multidisciplinary panels, not arbitrary, closed decision-making.
Consider cost-effectiveness	Consistent methods of assessing cost-effectiveness still need to be developed. But given the reality of resource constraints, evaluating value for money may be the best way to assure that highly effective treatments are available to all. Marginally effective treatments may assume lower priority.
Consider conditional device approvals	Same as in Table 4.
Consider covering clinical use of a device or technology only when application conditions are consistent with those under which it was shown to be effective and safe.	Same as in Table 3.

technologies be made publicly available at websites such as Clinicaltrials.gov shortly after their completion.⁵²

Malchau has proposed a stepwise approach to introducing new technologies, drawing on his experience with total hip arthroplasty.⁵³ This algorithm begins where the current regulatory process ends, and advances through stages of clinically relevant RCTs to postmarketing registry studies. Weinstein has proposed an alternative structure for funding clinical trials in the United States, based on a National Clinical Trials Consortium. This would distribute the burden of funding across the various stakeholders and insure independence in the conduct of trials, reducing conflicts of interest.⁵⁴

Angell recently called for the creation of an Institute for Prescription Drug Trials within the National Institutes of Health, to administer independent clinical trials of prescription drugs, including premarketing trials for drugs under FDA evaluation.⁵⁵ A similar model might apply to clinical research on spinal devices and technologies.

Congress is considering a major initiative in comparative effectiveness research which addresses some of the concerns addressed here, and for which the spine community should prepare.⁵⁶ The US Senate is currently evaluating legislation that would create a nonprofit Health Care Comparative Effectiveness Research Institute, with funding from both public and private sources. The new institute would supervise the creation of a national research agenda and commission independent research on the safety and effectiveness of a broad range of medical procedures.⁵⁷

Without changes such as those described in the following tables, our current course may lead to unnecessary and ineffective medical care, waste of limited resources, inappropriate restrictions on effective treatments, and impediments to important innovations. It will also continue to erode public confidence in spine care.

In current practice, once a device is approved by the FDA, the genie is out of the bottle, and hard to get back in. Obtaining the cooperation of physicians, participation by patients, and funding support from either government or industry to study a new technique at that point is almost impossible. We, therefore, favor implementation of more rigorous assessment before device approval, as well as more rigorous postmarketing surveillance. Innovation in the process of device development and evaluation—in addition to technical innovation itself—can make the spine field once again a leader in research and practice, rather than the mine field some now perceive it to be.

References

- Weinstein JN, Tosteson TD, Lurie JD, et al. Surgical vs. nonoperative treatment for lumbar disc herniation: the Spine Patient Outcomes Research Trial (SPORT): a randomized trial. *JAMA* 2006;296:2441–50.
- Weinstein JN, Lurie JD, Tosteson TD, et al. Surgical vs. nonoperative treatment for lumbar disc herniation: the Spine Patient Outcomes Research Trial (SPORT) observational cohort. *Spine* 2006;29:2451–59.
- Peul WC. Surgery vs. prolonged conservative treatment for sciatica. *N Engl J Med* 2007;356:2245–56.
- Weinstein JN, Tosteson TD, Lurie JD, et al. Surgical vs. nonsurgical therapy for lumbar spinal stenosis. *N Engl J Med* 2008;358:794–810.
- Weinstein JN, Lurie JD, Tosteson TD, et al. Surgical vs. nonsurgical treatment for lumbar degenerative spondylolisthesis. *Spine* 2007;35:2257–70.

6. Weber H. Lumbar disc herniation. A controlled, prospective study with ten years of observation. *Spine* 1983;8:131-40.
7. Weinstein JN, Tosteson TD, Lurie JD, et al. Surgical vs. nonoperative treatment for lumbar disc herniation: the Spine Patient Outcomes Research Trial (SPORT): a randomized trial. *JAMA* 2006;296:2441-50.
8. Peul WC, van den Hout WB, Brand R, et al. Prolonged conservative care versus early surgery in patients with sciatica caused by lumbar disc herniation: two-year results of a randomized controlled trial. *BMJ* 2008;336:1355-9.
9. Osterman H, Seitsalo S, Karpainen J, et al. Effectiveness of microdiscectomy for lumbar disc herniation: a randomized controlled trial with 2 years of follow-up. *Spine* 2006;31:2409-14.
10. Weinstein JN, Lurie JD, Olson PR, et al. United States trends and regional variations in lumbar spine surgery: 1992-2003. *Spine* 2006;27:707-14.
11. Deyo RA, Gray DT, Kreuter W, et al. United States trend in lumbar fusion surgery for degenerative conditions. *Spine* 2005;30:1441-45.
12. Abelson R. Medtronic will settle accusations on kickbacks. *New York Times*. July 19, 2006. Available at: <http://www.nytimes.com/2006/07/19/business/19medtronic.html>.
13. Abelson R. The spine as profit center. *New York Times*. December 30, 2006. Available at: <http://www.nytimes.com/2006/12/30/business/30spine.html>.
14. Consumer Reports. Ten overused medical tests and treatments. *Consumer Reports*. November 2007. Available at: <http://www.consumerreports.org/cro/cu-press-room/pressroom/archive/2007/11/0711-eng0711mtt.htm?resultPageIndex=1&resultIndex=2&searchTerm=10%20overused%20medical%20tests>.
15. Feigl DW, Gardner SN, McClellan M. Ensuring safe and effective medical devices. *Spine* 2003;34:191-2.
16. Deyo RA, Patrick DL. *Hope or Hype: The Obsession With Medical Advances and the High Cost of False Promises*. New York, NY: AMACOM; 2005:60-1, 163-4, 248-50.
17. Centers for Medicare and Medicaid Services. Proposed decision memo for lumbar artificial disc replacement (LADR) (CAG-00292R). May 25, 2007. Available at: www1.cms.hhs.gov/mcd/viewdraftdecisionmemo.asp?id=197.
18. Blumenthal S, McAfee PC, Guyer RD, et al. A prospective, randomized multicenter Food and Drug Administration investigational device exemptions study of lumbar total disc replacement with the Charité artificial disc versus lumbar fusion. *Spine* 2005;30:1565-75.
19. Shufflebarger HL. Surgeons, societies, and companies: ethics and legalities. *Spine* 2001;26:1947-49.
20. Hilibrand AS, Robbins M. Adjacent segment degeneration and adjacent segment disease: the consequences of spinal fusion? *Spine J* 2004;4(6 suppl):190S-4S.
21. FDA. Executive summary: report of US clinical study results (GO10188). Prestige Cervical Disc System. Available at: http://www.fda.gov/ohrms/dockets/ac/06/briefing/2006-4243b1_02.pdf. Accessed July 11, 2008.
22. FDA. Executive summary PO60023 Bryan Cervical Disc. Available at: <http://www.fda.gov/ohrms/dockets/ac/07/briefing/2007-4305b1-09.pdf>. Accessed July 11, 2008.
23. Shuchman M. Commercializing clinical trials—risks and benefits of the CRO boom. *N Engl J Med* 2007;357:1365-8.
24. Shah RV, Albert TJ, Bruegal-Sanchez V, et al. Industry support and correlation to study outcome for papers published in Spine. *Spine* 2005;30:1099-104; discussion 1105.
25. Scherer RW, Langenberg P, von Elm E. Full publication of results initially presented in abstracts. *Cochrane Database Syst Rev* 2007;18:MR000005.
26. Szpalski M, Gunzburg R, de Kleuver M. Unethical research funding contracts: just say NO! *Eur Spine J* 2003;12:107.
27. McAfee PC, Fedder IL, Saiedy S, et al. Experimental design of total disk replacement—experience with a prospective randomized study of the SB Charité. *Spine* 2003;28:S153-62.
28. McAfee PC, Fedder IL, Saiedy S, et al. SB Charité disc replacement: report of 60 prospective randomized cases in a US center. *J Spinal Disord Tech* 2003;16:424-33.
29. Guyer RD, McAfee PC, Hochschuler SH, et al. Prospective randomized study of the Charité artificial disc: data from two investigational centers. *Spine J* 2004;4(6 suppl):252S-9S.
30. Geisler FH, Blumenthal SL, Guyer RD, et al. Neurological complications of lumbar artificial disc replacement and comparison of clinical results with those related to lumbar arthrodesis in the literature: results of a multicenter, prospective, randomized investigational device exemption study of Charité intervertebral disc. Invited submission from the Joint Section Meeting on Disorders of the Spine and Peripheral Nerves, March 2004. *J Neurosurg Spine* 2004;1:143-54.
31. Blumenthal S, McAfee PC, Guyer RD, et al. A prospective, randomized, multicenter Food and Drug Administration investigational device exemption study of lumbar total disc replacement with the CHARITÉ artificial disc versus lumbar fusion, part I: evaluation of clinical outcomes. *Spine* 2005;30:1565-75.
32. McAfee PC, Cunningham B, Holsapple G, et al. A prospective, randomized, multicenter Food and Drug Administration investigational device exemption study of lumbar total disc replacement with the CHARITÉ artificial disc versus lumbar fusion, part II: evaluation of radiographic outcomes and correlation of surgical technique accuracy with clinical outcomes. *Spine* 2005;30:1576-83.
33. McAfee PC, Geisler FH, Saiedy SS, et al. Revisability of the CHARITÉ artificial disc replacement: analysis of 688 patients enrolled in the US. IDE study of the CHARITÉ Artificial Disc. *Spine* 2006;31:1217-26.
34. FDA. Summary of safety and effectiveness. Charité Artificial Disc. Available at: www.fda.gov/cdrh/PDF4/p040006b.pdf. Accessed July 11, 2008.
35. Sasso R, Smucker JD, Hacker RJ, et al. Artificial disc versus fusion: a prospective, randomized study with 2-year follow-up on 99 patients. *Spine* 2007;32:2933-40.
36. Anderson PA, Sasso RC, Riew KD, et al. Comparison of adverse events between the Bryan artificial cervical disc and anterior cervical arthrodesis. *Spine* 2008;33:1305-12.
37. Bertagnoli R, Yue JJ, Pfeiffer F, et al. Early results after ProDisc-C cervical disc replacement. *J Neurosurg Spine* 2005;2:403-10.
38. Bertagnoli R, Yue JJ, Shah RV, et al. The treatment of disabling multilevel lumbar discogenic low back pain with total disc arthroplasty utilizing the ProDisc prosthesis: a prospective study with 2-year minimum follow-up. *Spine* 2005;30:2192-9.
39. Blumenthal S, McAfee PC, Guyer RD, et al. A prospective, randomized, multicenter Food and Drug Administration investigational device exemptions study of lumbar total disc replacement with the CHARITÉ artificial disc versus lumbar fusion, part I: evaluation of clinical outcomes. *Spine* 2005;30:1565-75.
40. Zigler J, Delamarter R, Spivak JM, et al. Results of the prospective, randomized, multicenter Food and Drug Administration investigational device exemption study of the ProDisc-L total disc replacement versus circumferential fusion for the treatment of 1-level degenerative disc disease. *Spine* 2007;32:1155-62.
41. Baty P. When access to data is a real bone of contention. *Times Higher Education*. November 23, 2005. Available at: <http://www.timeshighereducation.co.uk/story.asp?sectioncode=26&storycode=199922>.
42. Blumsohn A. Authorship, ghost-science, access to data, and control of the pharmaceutical scientific literature: who stands behind the word? *AAAS Professional Ethics Report* 2006;19:1-4.
43. Ausman JI. I told you it was going to happen . . . *Surg Neurol* 2004;61:313-4.
44. Haldeman S, Dagenais S. A supermarket approach to the evidence-informed management of chronic low back pain. *Spine J* 2008;8:1-7.
45. Centers for Medicare and Medicaid Services. Lumbar artificial disc replacement. Pub 100-04. Medicare Claims Processing, September 21, 2007. Available at: <http://www.cms.hhs.gov/Transmittals/downloads/R1340CP.pdf>.
46. Centers for Medicare and Medicaid Services. Proposed decision memo for thermal intradiscal procedures. July 15, 2008. Available at: www.cms.hhs.gov/mcd/viewdraftdecisionmemo.asp?from2=viewdraftdecisionmemo.asp&id=2158&.
47. Angell M. Industry-sponsored clinical research: a broken system. *JAMA* 2008;300:1069-71.
48. Orszag PR, Ellis P. Addressing rising health care costs—a view from the Congressional Budget Office. *N Engl J Med* 2007;357:1885-7.
49. Brand RA, Buckwalter JA, Wright TM, et al. Patient care, professionalism, and relations with industry. *Clin Orthop Rel Res* 2008;466:517-9.
50. Drazen JM. Open clinical trials. *N Engl J Med* 2007;357:1756-7.
51. United States Code. US Public Law 110-85: Food and Drug Administration Amendments Act of 2007. 2008. Available at: http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=110_cong_public_laws&docid=f:publ085.110.
52. PLoS Medicine Editors. Next stop, don't block the doors: opening up access to clinical trial results. *PLoS Med* 2008;5:1007-8.
53. Malchau H. Introducing new technology: a stepwise algorithm. *Spine* 2000;25:285.
54. Weinstein JN. An altruistic approach to clinical trials: the National Clinical Trials Consortium (NCTC). *Spine* 2006;31:1-3.
55. Angell M. Industry-sponsored clinical research: a broken system. *JAMA* 2008;300:1069-71.
56. Ellis P et al. Research on the comparative effectiveness of medical treatments, issues, and options for an expanded federal role. Congressional Budget Office. December, 2007. Available at: <http://www.cbo.gov/ftpdocs/88xx/doc8891/088911.Frontmatter.1.2.shtml>.
57. Baucus M, Konrad K. Comparative Effectiveness Research Act of 2008.

- August 1, 2008. Available at: <http://finance.senate.gov/sitepages/leg/LEG%202008/080108%20CE%20Section-by-Section.pdf>.
58. Urrútia G, Kovacs F, Nishishinya MB, et al. Percutaneous thermocoagulation Intradiscal techniques for discogenic low back pain. *Spine* 2007;32:1146–54.
 59. Revel M, Payan C, Vallee C, et al. Automated percutaneous lumbar discectomy versus chemonucleolysis in the treatment of sciatica. A randomized multicenter trial. *Spine* 1993;18:1–7.
 60. Gibson AN, Waddell G. Surgical interventions for lumbar disc prolapse: updated Cochrane Review. *Spine* 2007;32:1735–47.
 61. Mezger R. Local firm's failure scars biotech field. *The Plain Dealer*. May 26, 2002. Available at: <http://www.cleveland.com/indepth/biotech/index.ssf?indepth/biotech/more/1022405449233330.html>.
 62. Fenton JJ, Mirza SK, Lahad A, et al. Variation in reported safety of lumbar interbody fusion: influence of industrial sponsorship and other study characteristics. *Spine* 2007;32:471–80.
 63. Shufflebarger HL. Surgeons, societies, and companies: ethics and legalities. *Spine* 26:1947–49.
 64. Jurecki-Tiller M, Bruening W, Tregear S, et al. Decompression therapy for the treatment of lumbosacral pain. Agency for Healthcare Research and Quality. April 26, 2007. Available at: <http://www.cms.hhs.gov/determinationprocess/downloads/id47TA.pdf>.
 65. Karmela Kroleza-Jeric, An-Wen Chan, Kay Dickersin, et al. Christian Gluud for the Ottawa Group. Principles for international registration of protocol information and results from human trials of health related interventions: Ottawa statement (part 1). *BMJ* 2005;330:956–8.
 66. World Health Organization international clinical trials registry platform. Available at: <http://www.who.int/ictrp/en/>.