

ADVERTIMENT. L'accés als continguts d'aquesta tesi queda condicionat a l'acceptació de les condicions d'ús establertes per la següent llicència Creative Commons:  <https://creativecommons.org/licenses/?lang=ca>

ADVERTENCIA. El acceso a los contenidos de esta tesis queda condicionado a la aceptación de las condiciones de uso establecidas por la siguiente licencia Creative Commons:  <https://creativecommons.org/licenses/?lang=es>

WARNING. The access to the contents of this doctoral thesis it is limited to the acceptance of the use conditions set by the following Creative Commons license:  <https://creativecommons.org/licenses/?lang=en>

TESIS DOCTORAL

PRAGMATISMO Y EFECTIVIDAD EN LOS ENSAYOS CLÍNICOS DE TERAPIA MANUAL:
UN ANÁLISIS METODOLÓGICO

Facultad de Medicina

Departamento de Pediatría, de Obstetricia y Ginecología, y de Medicina Preventiva y
Salud Pública

Programa de Doctorado en Metodología de la Investigación Biomédica y Salud Pública

Sònia Roura Carvajal

Directores: Dr. Gerard Urrútia Cuchí, Dra. Mercè Sitjà Rabert, Dr. Gerard Álvarez Bustins
Julio 2025

Tutor: Dr. Gerard Urrútia Cuchí



AGRADECIMIENTOS

A mis hijos Luka y Giulia, a Paco y a mi familia, por las horas que les he quitado y las horas que han destinado a ayudarme.

A mis directores Gerard Urrútia i Mercè Sitjà y al coordinador del programa Xavier Bonfill por su apoyo incondicional durante todo el proceso.

A mis co-autores en las publicaciones de este trabajo de tesis, en especial a Iván Solá, Javier Brachiglione y Rodrigo Núñez-Cortés quienes me han acompañado des de los inicios del proyecto en el Máster.

A mi director, compañero y amigo Gerard Álvarez quién sin él esta tesis no habría sido posible.

“Sólo aquellos que se atreven a ir demasiado lejos pueden descubrir hasta dónde pueden llegar.”

T. S. Eliot

RESUMEN

Antecedentes

La Terapia Manual es un abordaje que emplea movimientos pasivos aplicados por el profesional para tratar dolor y limitación funcional en afecciones musculoesqueléticas. Se basa en razonamiento clínico, diagnóstico y un enfoque centrado en la persona, resaltando la interacción paciente-terapeuta. Sus efectos se atribuyen a mecanismos neuromoduladores, aunque su relación con los resultados clínicos requiere más investigación. Al tratarse de una intervención compleja donde varios componentes intervienen en ella (tacto, interacción paciente-terapeuta, efectos placebo, nocebo y factores contextuales, medidas de resultado auto-reportadas por el paciente, individualización del tratamiento) su evaluación a nivel de investigación presenta retos metodológicos considerables (protocolizar diseños, controlar parámetros de ejecución, homogeneizar poblaciones y definir controles placebo o sham). Los estudios pragmáticos evalúan la efectividad en condiciones reales y se han postulado como un tipo de diseño que se puede adaptar a las particularidades de la terapia manual. Es por ello que resulta importante conocer su uso entre la comunidad científica y evaluar su idoneidad para tal propósito.

Objetivos

Este trabajo de tesis tiene como objetivo evaluar de forma exhaustiva las características de los estudios de efectividad publicados en el campo de la terapia manual (pragmatismo, reporte y riesgo de sesgo), tanto de aquellos estudios que se autodenominan pragmáticos como los que no. Más allá de que los autores pretendan estudiar los efectos de una intervención en condiciones parecidas a la práctica real, se pretende evaluar si el diseño metodológico de los estudios es acorde a la pregunta de investigación planteada en los estudios.

Métodos

Se han llevado a cabo 3 trabajos de revisión a partir de una amplia muestra de estudios de efectividad en terapia manual. En los dos primeros trabajos se analizaron todos los

estudios de efectividad que se autodenominaban pragmáticos (en relación a su diseño) en su título y/o resumen, publicados en Terapia Manual. Se evaluó el pragmatismo de los estudios a través de la herramienta PRECIS 2, su reporte a través de las guías CONSORT y su riesgo de sesgo a través de la herramienta Cochrane Risk of Bias, y se describieron sus características. En el tercer trabajo se amplió la muestra a estudios de efectividad no autodenominados pragmáticos y se analizaron sus características para compararlas posteriormente con las de los estudios autodenominados pragmáticos.

Resultados

Los resultados de los dos primeros trabajos muestran que a pesar de la creciente atención a los estudios pragmáticos, la literatura científica no muestra un aumento en este tipo de diseño aplicado a la Terapia Manual. Nuestros resultados sugieren que los estudios que se autodenominan pragmáticos en Terapia Manual presentan una actitud moderadamente pragmática: emplean características altamente explicativas, como el enmascaramiento, el entorno unicéntrico o un control placebo, lo que cuestiona la etiqueta «pragmático». Los autores no justifican suficientemente por qué su estudio busca ser pragmático ni detallan qué aspectos del diseño refuerzan ese pragmatismo. La calidad del reporte de los estudios pragmáticos en Terapia Manual es subóptima y se observa la presencia frecuente de sesgos de realización y de detección, ligados a las dificultades intrínsecas de este tipo de intervenciones para mantener el enmascaramiento, y un sesgo de reporte en la mitad de los casos. Asimismo, encontramos una clara correlación entre la calidad del reporte y el riesgo de sesgo de los estudios.

Los resultados del tercer trabajo mostraron que el número de ensayos clínicos aleatorizados de efectividad en Terapia Manual ha aumentado con el tiempo y, en general, muestran una actitud moderadamente pragmática, independientemente de si se autodenominan pragmáticos o del uso de un grupo de control con o sin placebo. A pesar de que la mayoría pretende emular la práctica clínica, sus diseños no siguen de manera sistemática la metodología pragmática. El reporte de estos estudios es, en su mayoría, inadecuado, empeorando cuando no usan la etiqueta «pragmático» y cuando el control

es “usual care” u otra intervención activa en lugar de un placebo. El sesgo de realización está presente en prácticamente toda la muestra; el sesgo de selección es menor en los estudios autodenominados pragmáticos, y el de detección resulta más bajo cuando se emplea un control placebo que con otras intervenciones o “usual care”. Además, la distinción entre los términos de eficacia y efectividad no está clara en la comunidad investigadora de Terapia Manual, empleando a menudo el término efectividad en diseños con características explicativas.

Conclusiones

Nuestros resultados sugieren que los ensayos de terapia manual que se autodenominan “pragmáticos” no han incrementado su grado de pragmatismo a lo largo del tiempo, sino que mantienen una actitud moderada. Esto significa que, aunque reclamen emular la práctica clínica real, muchos siguen introduciendo elementos muy explicativos que alejan el diseño de la flexibilidad de la práctica real. Esto nos indica que la simple etiqueta “pragmático” no basta: es esencial justificar en qué medida cada decisión metodológica (criterios de inclusión, reclutamiento, tipo de intervención, análisis) favorece o limita la aplicabilidad real de los resultados.

Además, el uso indiscriminado del término “efectividad” sin atender a sus implicaciones pragmáticas demuestra la necesidad de insistir en la diferenciación entre eficacia y efectividad. En la práctica, nuestros estudios resultan pragmáticos en lo que respecta a la selección de participantes y a las medidas de resultado, pero siguen siendo explicativos cuando se trata de estandarizar demasiado la intervención o el entorno, aspectos que rara vez coinciden exactamente con la realidad clínica diaria.

En cuanto a la calidad del reporte, constatamos que permanece subóptima: tan solo los ensayos con bajo riesgo de sesgo cumplen con mayor fidelidad las guías CONSORT. Si queremos que la investigación en Terapia Manual sea útil para clínicos, se debe poner más esfuerzo en reportar con transparencia cada procedimiento (aleatorización, ocultación de la asignación, seguimiento de participantes, detalles precisos de la técnica aplicada). Solo así se mejora la reproducibilidad y la credibilidad de los resultados.

Finalmente, los sesgos de realización y detección, casi inevitables por las dificultades de

enmascarar una intervención física, ponen de manifiesto la importancia de reconocer explícitamente estas limitaciones y buscar alternativas (p. ej., evaluadores externos, medidas objetivas suplementarias) que ayuden a mitigar su impacto. En definitiva, estos resultados subrayan que lograr un verdadero pragmatismo en un ensayo de Terapia Manual exige: 1) construir un diseño que emule al máximo a la práctica clínica real; 2) emplear herramientas como PRECIS-2 para ayudar y evaluar el diseño; 3) cumplir rigurosamente con las guías de reporte; y 4) reconocer y manejar los sesgos inherentes. Solo así podremos generar evidencia que realmente informe y transforme la atención al paciente en contextos de práctica habitual.

RESUM

Antecedents

La Teràpia Manual és un abordatge que empra moviments passius aplicats pel professional per tractar dolor i limitació funcional, principalment, en afeccions músculo-esquelètiques. Es basa en realitzar raonament clínic, diagnòstic i un enfocament centrat en la persona, ressaltant la interacció pacient-terapeuta. Els seus efectes s'atribueixen a mecanismes neuro-moduladors, encara que la seva relació amb els resultats clínics requereix més investigació. En tractar-se d'una intervenció complexa on diversos components intervenen en ella (tacte, interacció pacient-terapeuta, efectes placebo, nocebo i factors contextuais, mesures de resultat auto-reportades pel pacient, individualització del tractament), la seva avaluació a nivell d'investigació presenta reptes metodològics considerables (protocol·litzar dissenys, controlar paràmetres d'execució, homogeneïtzar poblacions i definir controls placebo o sham). Els estudis pragmàtics avaluen l'efectivitat en condicions reals i s'han postulat com un tipus de disseny que es pot adaptar a les particularitats de la teràpia manual. És per això que resulta important conèixer el seu ús entre la comunitat científica i avaluar-ne la idoneïtat per a tal propòsit.

Objectius

Aquest treball de tesi té com a objectiu avaluar de forma exhaustiva les característiques dels estudis d'efectivitat publicats en el camp de la teràpia manual (pragmatisme, report i risc de biaix), tant d'aquells estudis que s'autodenominen pragmàtics com els que no. Més enllà que els autors pretenguin estudiar els efectes d'una intervenció en condicions semblants a la pràctica real, es pretén avaluar si el disseny metodològic dels estudis és acord amb la pregunta de recerca plantejada.

Mètodes

S'han dut a terme 3 treballs de revisió a partir d'una àmplia mostra d'estudis d'efectivitat en Teràpia Manual. En els dos primers treballs es van analitzar tots els estudis d'efectivitat

que s'autodenominaven pragmàtics (en relació amb el seu disseny) en el seu títol i/o resum publicats en Teràpia Manual. Es va avaluar el pragmatisme dels estudis a través de l'eina PRECIS 2, el seu report a través de les guies CONSORT i el seu risc de biaix a través de l'eina Cochrane Risk of Bias, i es van descriure les seves característiques. En el tercer treball es va ampliar la mostra a estudis d'efectivitat no autodenominats pragmàtics i es van analitzar les seves característiques per a comparar-les posteriorment amb les dels estudis autodenominats pragmàtics.

Resultats

Els resultats dels dos primers treballs mostren que, malgrat l'atenció creixent als estudis pragmàtics, la literatura científica no mostra un augment en aquest tipus de disseny aplicat a la Teràpia Manual. Els nostres resultats suggereixen que aquells que s'autodenominen pragmàtics presenten una actitud moderadament pragmàtica: emprant característiques altament explicatives, com l'emascament, l'entorn unicèntric o un control placebo, cosa que qüestiona l'etiqueta «pragmàtic». Els autors no justificaven prou per què el seu estudi pretenia ser pragmàtic ni detallaven quins aspectes del disseny en reforcen aquest pragmatisme. La qualitat del report dels assajos pragmàtics en Teràpia Manual és subòptima i s'observa la presència freqüent de biaixos de realització i detecció, lligats a les dificultats intrínseques d'aquest tipus d'intervenció per mantenir l'emascament, i un biaix de report en la meitat dels casos. Així mateix, vam trobar una clara correlació entre la qualitat del report i el risc de biaix dels estudis.

Els resultats del tercer treball han mostrat que el nombre d'assaigs clínics aleatoritzats d'efectivitat en teràpia manual ha augmentat amb el temps i, en general, mostren una actitud moderadament pragmàtica, independentment de si s'autodenominen pragmàtics o de l'ús d'un grup de control amb o sense placebo. Malgrat que la majoria pretenia emular la pràctica clínica, els seus dissenys no segueixen de manera sistemàtica la metodologia pragmàtica. L'informe d'aquests estudis és, en la seva majoria, inadequat, empitjorant quan no fan servir l'etiqueta «pragmàtic» i quan el control és “usual care” o una altra intervenció en lloc d'un placebo/sham. El biaix de realització està present en pràcticament tota la mostra; el biaix de selecció és menor en els estudis autodenominats

pragmàtics, i el de detecció resulta més baix quan s'empra un control placebo/sham que amb altres intervencions o "usual care". A més, la distinció entre eficàcia i efectivitat sembla no estar clara en la comunitat investigadora de Teràpia Manual, utilitzant sovint el terme efectivitat en dissenys amb característiques explicatives.

Conclusions

Els nostres resultats suggereixen que els assaigs de Teràpia Manual que s'autodenominen "pragmàtics" no han incrementat el seu grau de pragmatisme al llarg del temps, sinó que mantenen una actitud moderada. Això significa que, tot i que reivindiquin emular la pràctica clínica real, molts continuen introduint elements molt explicatius que allunyen el disseny de la flexibilitat de la pràctica real. Això ens indica que l'etiqueta "pragmàtic" per si sola no és suficient: és essencial justificar en quina mesura cada decisió metodològica (criteris d'inclusió, reclutament, tipus d'intervenció, anàlisi) afavoreix o limita l'aplicabilitat real dels resultats.

A més, l'ús indiscriminat del terme "efectivitat" sense atendre a les seves implicacions pragmàtiques demostra la necessitat d'insistir en la diferenciació entre eficàcia i efectivitat. A la pràctica, els nostres estudis resulten pragmàtics pel que fa a la selecció de participants i a les mesures de resultat (factors que reflecteixen la realitat del pacient), però continuen sent explicatius quan es tracta d'estandarditzar massa la intervenció o l'entorn, aspectes que rarament coincideixen exactament amb la realitat clínica diària.

Pel que fa a la qualitat del report, constatem que continua subòptima: només els assaigs amb baix risc de biaix compleixen amb més fidelitat les guies CONSORT. Si volem que la recerca en Teràpia Manual sigui útil per a clínics, cal posar més esforç en informar amb transparència cada procediment (randomització, ocultació de l'assignació, seguiment de participants, detalls precisos de la tècnica aplicada). Només així es millora la reproductibilitat i la credibilitat dels resultats.

Finalment, els biaixos de realització i detecció, gairebé inevitables per les dificultats d'emascarar una intervenció física, posen de manifest la importància de reconèixer de manera explícita aquestes limitacions i cercar alternatives (per exemple, avaluadors externs, mesures objectives suplementàries) que ajudin a mitigar-ne l'impacte. En

definitiva, aquests resultats subratllen que aconseguir un veritable pragmatisme en un assaig de teràpia manual exigeix: 1) construir un disseny que enllaci clarament cada decisió amb la pràctica clínica real; 2) emprar eines com PRECIS-2 per avaluar cada domini; 3) complir rigorosament amb les guies de report; i 4) reconèixer i gestionar els biaixos inherents. Només així podrem generar evidència que realment informi i transformi l'atenció al pacient en contextos clínics reals.

ABSTRACT

Background

Manual Therapy uses passive movements applied by a practitioner to address pain and functional limitations in musculoskeletal conditions. It relies on clinical reasoning, diagnosis, and a patient-centred approach, emphasising the patient–therapist interaction. Manual Therapy effects are thought to be neuromodulator, but the link to clinical outcomes requires further study. As a complex intervention (including touch, patient–therapist interaction, placebo and nocebo effects, contextual factors, patient-reported outcome measures, and treatment individualisation), its evaluation poses substantial methodological challenges (standardising protocols, controlling execution parameters, homogenising populations, and defining placebo or sham controls). Pragmatic trials, which assess effectiveness under real-world conditions, have been proposed as suitable designs for Manual Therapy; therefore, it is crucial to understand their use and appropriateness within the scientific community.

Objectives

This doctoral thesis aims to comprehensively evaluate the characteristics of published Manual therapy effectiveness trials, including both those self-labelled as pragmatic and those that are not, focusing on pragmatism, reporting quality, and risk of bias. Beyond merely claiming to study interventions under real-world conditions, the goal is to determine whether each trial’s methodological design aligns with its research question.

Methods

Three systematic reviews were conducted on a broad sample of trials evaluating the effectiveness of manual therapy. The first two reviews analysed all trials explicitly labelled “pragmatic” in their title and/or abstract. Each study’s pragmatism was assessed using the PRECIS-2 tool, reporting quality was evaluated according to CONSORT guidelines, and the risk of bias was assessed using the Cochrane Risk of Bias tool; their main

characteristics were described. In the third review, the sample was expanded to include effectiveness trials not labelled “pragmatic,” and their features were compared with those of the self-labelled pragmatic trials.

Results

In the first two reviews, despite growing interest in pragmatic Manual therapy trials, the literature does not show an increase in their use. Self-labelled pragmatic Manual therapy trials exhibit only a moderately pragmatic attitude, as many still incorporate highly explanatory elements, such as strict blinding, single-centre settings, or placebo controls, which calls into question their “pragmatic” label. Authors seldom justify why their trial is intended to be pragmatic or specify which design aspects reinforce that pragmatism. Reporting quality in these pragmatic trials is suboptimal, with frequent performance and detection biases, primarily due to inherent blinding challenges associated with this type of intervention and reporting bias in approximately half of the cases. A clear correlation was observed between reporting quality and risk of bias.

In the third review, the number of randomised clinical effectiveness trials in Manual therapy has increased over time; in general, they too show a moderately pragmatic stance, irrespective of whether they self-label as pragmatic or use placebo controls. Although most trials aim to emulate clinical practice, their designs do not systematically follow pragmatic methodology. Reporting remains largely inadequate, especially when the “pragmatic” label is absent and when “usual care” or another active intervention is used instead of a placebo/sham. Performance bias is nearly universal across the sample; selection bias is lower among self-labelled pragmatic trials, and detection bias decreases when using placebo controls rather than “usual care” or other active comparators. Additionally, the distinction between “efficacy” and “effectiveness” remains unclear in the Manual Therapy research community, with “effectiveness” often applied to designs that include explanatory characteristics.

Conclusions

Self-labelled pragmatic Manual therapy trials have not increased their level of pragmatism over time but maintain a moderate attitude. Although they claim to emulate real-world

practice, many continue to introduce highly explanatory elements that distance their design from everyday clinical flexibility. Thus, the “pragmatic” label alone is insufficient; researchers must justify exactly how each methodological choice (inclusion criteria, recruitment, intervention type, analysis) supports or limits real-world applicability. Moreover, the indiscriminate use of “effectiveness” without attending to its pragmatic implications underscores the need to clarify the efficacy–effectiveness distinction. In practice, Manual Therapy trials may be pragmatic regarding participant selection and outcome measures—reflecting patient reality—but remain explanatory in terms of intervention standardisation and setting, which seldom mirror actual clinical environments.

Regarding reporting quality, we found that it remains suboptimal: only trials with a low risk of bias closely follow the CONSORT guidelines. To ensure Manual Therapy research is useful to clinicians and decision makers, greater effort must be made to transparently report every procedure (randomisation, allocation concealment, participant follow-up, precise technique details). Only through such transparency can reproducibility and credibility improve.

Finally, performance and detection biases, almost inevitable given Manual Therapy interventions’ blinding challenges, highlight the importance of explicitly acknowledging these limitations and seeking alternatives (e.g., external assessors, supplementary objective measures) to mitigate their impact. Ultimately, achieving genuine pragmatism in a Manual Therapy trial requires: (1) designing each decision to align clearly with real clinical practice; (2) using tools like PRECIS-2 to evaluate every domain; (3) rigorously adhering to reporting guidelines; and (4) recognising and managing inherent biases. Only then can we generate evidence that genuinely informs and improves patient care in everyday contexts.

ÍNDICE

1. INTRODUCCIÓN	5
1.1 Terapia Manual e Intervenciones Complejas	5
1.1.1 La Terapia Manual como intervención compleja.....	5
1.1.2 Características particulares de la Terapia Manual.....	8
1.2 INVESTIGACIÓN EN TERAPIA MANUAL.....	14
1.2.1 Estado de la investigación en Terapia Manual.....	14
1.2.2 El ensayo clínico aleatorizado en Terapia Manual.....	15
1.2.3 Dificultades metodológicas de la investigación en Terapia Manual	18
1.2.4 Propuestas metodológicas adaptadas a la Terapia Manual.....	24
1.3 ENSAYOS CLÍNICOS ALEATORIZADOS PRAGMÁTICOS	32
1.3.1 Diseño de ensayos clínicos aleatorizados pragmáticos	33
1.3.2 Evaluación retrospectiva de la actitud pragmática de los ECAs.....	36
1.3.3 Limitaciones de los ECAs pragmáticos	36
1.4 JUSTIFICACIÓN DE LA TESIS.....	39
2. OBJETIVOS	42
2.1 Objetivo principal.....	42
2.2 Objetivos secundarios	42
3. MÉTODOS	44
3.1 Métodos del primer trabajo	44
3.1.1 Criterios de elegibilidad de los estudios incluidos.....	44
3.1.2 Estrategia de búsqueda.....	45
3.1.3 Selección de los estudios	45
3.1.4 Proceso de recopilación y extracción de datos.....	45
3.1.5 Análisis de datos	48
3.2 Métodos del segundo trabajo.....	49
3.2.1 Criterios de elegibilidad de los estudios incluidos.....	49
3.2.2 Estrategia de búsqueda.....	50
3.2.3 Selección de los estudios	50

3.2.4	Proceso de recopilación y extracción de datos.....	50
3.2.5	Análisis de datos	50
3.3	Métodos del tercer trabajo	50
3.3.1	Criterios de elegibilidad de los estudios incluidos.....	51
3.3.2	Estrategia de búsqueda.....	51
3.3.3	Selección de los estudios	51
3.3.4	Proceso de recopilación y extracción de datos.....	52
3.3.5	Análisis de datos	52
4.	RESULTADOS	54
4.1	Resultados del primer trabajo.....	55
4.1.1	Resultados de la búsqueda	55
4.1.2	Características principales de los estudios autodenominados pragmáticos	55
4.1.3	Evaluación retrospectiva del pragmatismo con la herramienta PRECIS-2.....	56
4.1.4	Primera publicación	57
4.2	Resultados del segundo trabajo	70
4.2.1	Evaluación del reporte a través del cumplimiento de los ítems de la guía CONSORT	70
4.2.2	Evaluación del riesgo de sesgo.....	71
4.2.3	Relación entre el pragmatismo, la calidad del reporte y el riesgo de sesgo	71
4.2.4	Segunda publicación	72
4.3	Resultados del tercer trabajo	83
4.3.1	Resultados de la búsqueda	83
4.3.2	Comparación de las características principales de los estudios incluidos.....	83
4.3.3	Evaluación retrospectiva del pragmatismo a través de la herramienta PRECIS-2	84
4.3.4	Evaluación del reporte a través del cumplimiento de los ítems de la guía CONSORT	85
4.3.5	Evaluación del Riesgo de sesgo	86
4.3.6	Tercera publicación	87
5.	DISCUSIÓN.....	100
5.1	Principales resultados derivados de los trabajos	100
5.2	Discusión de los elementos más relevantes de los estudios de efectividad.	102
5.2.1	Estado de la investigación sobre la efectividad de la Terapia Manual	102
5.2.2	Actitud pragmática y características metodológicas de los ECAs de efectividad en Terapia Manual	103

5.2.3 Reporte de los estudios de efectividad	113
5.2.4 Riesgo de sesgo de los estudios de efectividad.....	116
5.4.5 Relación pragmatismo-reporte-riesgo de sesgo	119
5.3 Idoneidad en el uso de la terminología eficacia/efectividad en los estudios de TM.....	120
5.4. Fortalezas, limitaciones e implicaciones.....	121
5.4.1 Fortalezas.....	122
5.4.2 Limitaciones	122
5.4.3 Implicaciones.....	123
6. CONCLUSIONES	142
7. ABREVIATURAS	143
8. ANEXOS	145
8.1 Anexo 1	145
8.2 Anexo 2	149
8.3 Anexo 3	159
9. BIBLIOGRAFÍA	194

1. INTRODUCCIÓN

1.1 Terapia Manual e Intervenciones Complejas

La Terapia Manual (TM) es un abordaje terapéutico que consiste en la aplicación de movimientos manuales pasivos con intención terapéutica (realizados por el terapeuta), utilizado para tratar diferentes tipos de problemas de salud, aunque su actuación está especialmente reconocida en el tratamiento del dolor y la impotencia funcional, sobre todo debido a afecciones músculo-esqueléticas [1–7]. Se basa en un proceso de razonamiento clínico y diagnóstico, con un enfoque centrado en la persona y en el que la interacción paciente/terapeuta es muy importante, debido a los múltiples factores que pueden intervenir en esta interacción [4,8]. Abarca una amplia gama de intervenciones, como técnicas de tejidos blandos, movilizaciones o manipulaciones articulares, masaje, tacto ligero, liberación miofascial, manipulación nerviosa, tensión/contratensión y acupresión [1,6]. La TM es utilizada por una gran variedad de profesiones en el ámbito de la salud (fisioterapia, osteopatía, quiropráctica, masoterapia, entre otras) y desempeña un papel relevante en la atención sanitaria tanto pública como privada [5,7,9]. Sus mecanismos de actuación se han estudiado ampliamente en la literatura [10,11], y si bien actualmente se postulan bajo un efecto neuromodulador [4,12–15], la correlación entre estos mecanismos y los resultados clínicos no se ha establecido de forma clara [6]. Es por este motivo que Keter et al. han publicado recientemente un documento de consenso de expertos internacional señalando qué ítems relacionados con los mecanismos de acción de la TM deberían ser investigados en más profundidad [6]. Entre ellos, los mecanismos neurológicos, neuro-inmunes, biomecánicos, traslacionales y factores contextuales [6].

1.1.1 La Terapia Manual como intervención compleja

Las intervenciones complejas (IC), a diferencia de las intervenciones farmacológicas simples, son aquellas intervenciones que están formadas por distintos componentes que interactúan entre sí [16]. El Medical Research Council ha definido la IC en su guía del siguiente modo: "Una intervención puede considerarse compleja debido a las propiedades de la propia intervención, como el número de componentes implicados; la gama de comportamientos a los que se dirige; la experiencia por parte de quienes realizan y reciben la intervención; el número de grupos, entornos o niveles a los que se dirige; o el nivel permitido de flexibilidad de la intervención o de sus componentes" [16]. Este documento explica que "cuánto más difícil sea definir exactamente cuáles son los ingredientes activos de la intervención y cómo se relacionan entre ellos, mayor es la probabilidad de que se trate de una intervención compleja" [17]. Las intervenciones de TM se consideran ICs debido a la interacción de múltiples factores que van más allá de la simple aplicación de una técnica [12,13,18]. Esta complejidad viene dada principalmente por: la TM comporta la utilización de diversas técnicas y/o abordajes terapéuticos que se combinan para conseguir un resultado clínico [1,13]; además, los efectos de la TM pueden estar influenciados por la destreza terapéutica de cada profesional que las aplica, la intensidad y frecuencia de las intervenciones, los factores contextuales que pueden interferir y/o complementar la intervención, las expectativas del paciente y sus vivencias y el efecto placebo, entre otros [4,19–22]. Esta complejidad puede comportar dificultades para entender y evaluar el mecanismo de acción que existe detrás de la intervención y saber si una intervención es efectiva en la práctica real [13,20–23]. Las características particulares de la TM como IC están descritas en la **Figura 1** y desarrolladas a continuación para profundizar en sus elementos clave y ver en qué difieren y de qué modo con respecto a otras intervenciones.

Figura 1. Características de la Terapia Manual como Intervención Compleja (tabla adaptada de Craig et al. 2008) [24]



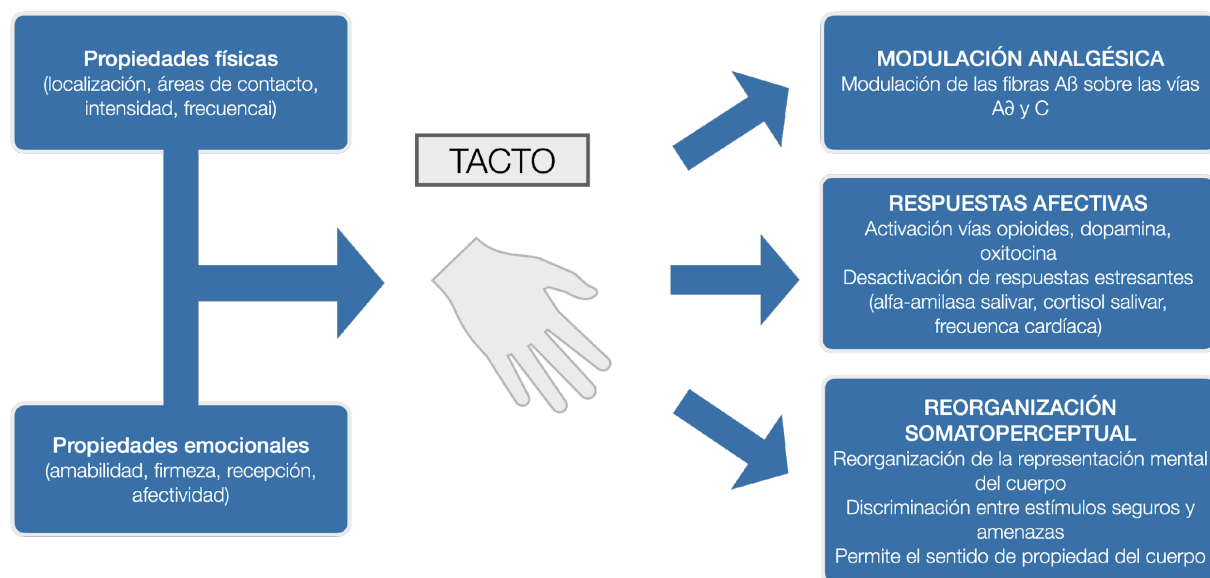
1.1.2 Características particulares de la Terapia Manual

Todas las características que hacen de la terapia manual una intervención compleja pueden agruparse en los siguientes puntos:

1) Tacto como herramienta de diagnóstico y tratamiento

La principal característica que diferencia la TM de otros campos es el uso del tacto en sus intervenciones (tanto para el diagnóstico como para el tratamiento) [7]. El uso del tacto, conocido actualmente como técnicas “hands on”, tanto en la intervención experimental como en la intervención control, contiene elementos que pueden no ser fáciles de controlar durante un estudio de investigación [18,25]. Estos elementos incluyen el tipo de contacto que se aplica, la cantidad de presión que se ejerce, el tipo de movimiento que se realiza y la posición del paciente, las áreas donde se contacta o el tiempo que se contacta. Estas particularidades pueden hacer que reproducir la misma técnica por un mismo terapeuta o por otros terapeutas sea muy complicado [25]. Además, la experiencia y la destreza del terapeuta pueden influir también sobre el tacto como intervención y, consecuentemente, modificar la efectividad de la técnica [13,25,26]. La literatura científica identifica las técnicas “hands on” como elementos moduladores del dolor a través de diferentes hipótesis. Actualmente, los efectos neurofisiológicos y los efectos contextuales parecen ser las hipótesis con mayor peso para poder explicar la capacidad de las técnicas manuales para ayudar a modular cómo el paciente procesa el dolor [4,13–15,27,28]. Geri et al. ofrecen un cambio de perspectiva planteando la TM como una forma de comunicación con el paciente de forma no verbal, capaz de producir analgesia, regular las emociones del paciente y reorganizar las representaciones mentales [27] (Figura 2).

Figura 2. Influencia del tacto en los tratamientos con Terapia Manual (adaptada con permiso de Geri et al) [27]



2) Interacción paciente-terapeuta

La compleja interacción entre paciente y terapeuta abarca diferentes dimensiones que pueden ser difíciles de controlar en el contexto de una investigación (**Figura 3**):

1) Uso del tacto: como se ha explicado anteriormente, el uso del tacto como intervención puede influenciar la respuesta del paciente.

2) Efectos placebo y nocebo: el efecto placebo es aquel efecto terapéutico conseguido por un contexto asistencial positivo que puede mejorar los síntomas del paciente, al contrario que el efecto nocebo donde se producen efectos de empeoramiento de los síntomas por un contexto negativo [29]. Actualmente, estos efectos se consideran componentes activos de los mecanismos de alivio del dolor a través de la TM [12,30]. Teniendo en cuenta los mecanismos neurofisiológicos de la TM, el placebo y nocebo

constituyen factores importantes a tener en cuenta en los mecanismos supraespinales top-down como moduladores de la sintomatología del paciente [29].

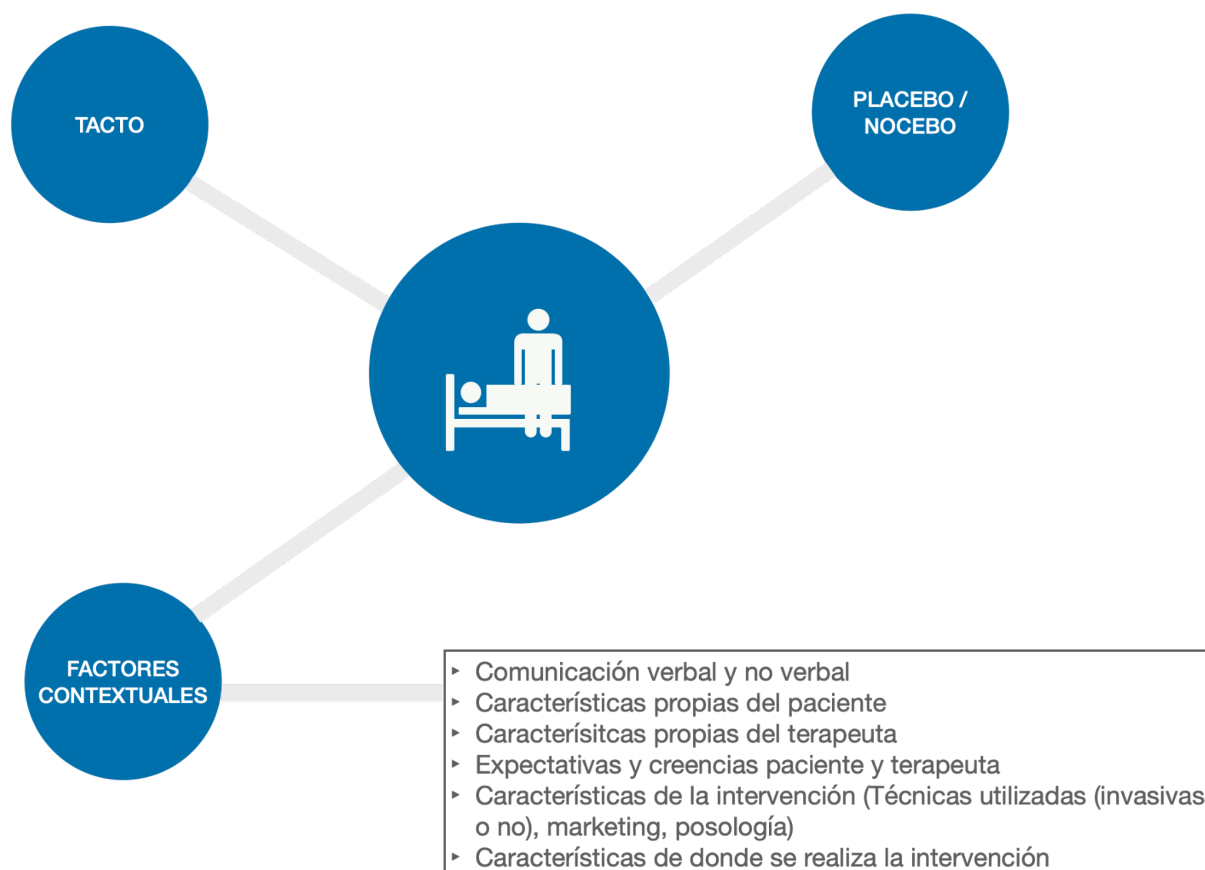
3) Factores contextuales intrínsecos y extrínsecos: Son los factores relacionados con el modo cómo se aplican las intervenciones y son otro punto importante a tener en cuenta durante el encuentro paciente-terapeuta [12,29–31]. Estos factores contextuales son (tabla 1):

Tabla 1. Factores contextuales en el encuentro paciente-terapeuta

Factores contextuales en el encuentro paciente-terapeuta		
Características del terapeuta		Profesionalidad: experiencia, cualificación, reputación, formación. Mentalidad: comportamiento, creencias, expectativas, experiencia previa. Aspecto: presencia, uniforme, confianza que muestra.
Características del paciente		Mentalidad: expectativas, experiencia previa, historial de tratamiento, preferencias. Situación de partida: nivel de síntomas, comorbilidad, estado de salud, sexo, edad.
Relación terapeuta-paciente		Comunicación verbal: mensaje positivo, tono de voz, escucha activa, sugerencias de apoyo y ánimo, reciprocidad del lenguaje, calidez, atención, cuidado, interacción empática. Comunicación no verbal: contacto visual, expresión facial agradable, sonrisa, postura, gestos, asentimiento con la cabeza, inclinación hacia delante, orientación corporal abierta.
Características del tratamiento		Tacto terapéutico: emocional, empático, afectivo Modalidad: nivel de invasividad Posología: trato personalizado, tratamiento impartido por

	<p>el mismo fisioterapeuta, limpieza, duración adecuada de la consulta, puntualidad, flexibilidad con las citas del paciente, tratamiento puntual y eficaz, frecuencia, duración y seguimiento adecuados de la terapia</p> <p>Marketing: marca, premio, novedad, ritualidad.</p>
Características del entorno sanitario	<p>Distractores positivos: iluminación natural, bajos niveles de ruido, música relajante y suave, aromas agradables, temperatura adecuada</p> <p>indicaciones de apoyo: señales muy visibles y fáciles de leer, información sobre aparcamientos entradas accesibles, indicaciones verbales o escritas claras y coherentes, mostradores de información e información electrónica accesible.</p> <p>Elementos de confort: ventanas y claraboyas, entornos terapéuticos privados, buen acceso a los servicios, horarios convenientes, ubicación, aparcamiento y personal de apoyo disponible y accesible, decoraciones y ornamentos (obras de arte de la naturaleza, vegetación verde, flores, agua, plantas, jardín, color).</p>

Figura 3: Dimensiones de la interacción paciente-terapeuta



3) Medidas de resultado

La medición de los efectos del tratamiento en la investigación sobre la TM difiere de la investigación biomédica en varios aspectos ya que, en muchas ocasiones, no existe una relación lineal entre la intervención y el efecto deseado [18]. En TM es particularmente importante utilizar variables relevantes y significativas para los pacientes, donde se evalúe, aparte de alguna variable fisiológica, la funcionalidad y la afectación en las actividades de la vida diaria y/o social, entre otras [18,21,32,33]. Este tipo de medidas se denominan “*patient-reported outcomes measures*” (PROMs) donde se puede evaluar diferentes variables de resultado desde la perspectiva del paciente [33]. Estas variables se pueden medir a través de respuestas informales, escalas de dolor, cuestionarios de

discapacidad y/o satisfacción validados, donde, a través de unos algoritmos, se consigue cuantificar la respuesta de manera objetiva [5,18,21,32,34]. Para constructos como el “dolor” o la “satisfacción” se debería considerar el uso de PROMs como el “Gold standard” [34]. El dolor, la movilidad y la calidad de vida son resultados muy subjetivos de cada paciente y usando medidas validadas se pueden obtener resultados muy objetivos [34]. Lo importante es escoger una variable realista y razonable que realmente dé respuesta a la pregunta de investigación [18]. No obstante, su análisis estadístico es difícil de comparar entre pacientes y se necesita más investigación para asegurar esta comparación [35,36].

4) Técnicas, abordajes e Individualización del tratamiento (patient-centred)

La TM engloba muchas técnicas diferentes que pueden ser usadas de forma aislada o en combinación entre ellas [5]. Las diferentes técnicas incluyen:

- Técnicas de tejidos blandos: Movilización de tejidos, como músculos, tendones o ligamentos, sin provocar movimiento o cambio de posición articular, por ejemplo, masaje, técnica de energía muscular, liberación miofascial o de puntos gatillo.
- Tracción: Distracción manual de una parte del cuerpo, por ejemplo, el cuello.
- Manipulación y movilización: Técnicas manuales aplicadas específicamente a las articulaciones. La manipulación es la aplicación de una fuerza de alta velocidad y baja amplitud cerca del final de la amplitud de las articulaciones. A menudo, pero no siempre, va acompañada de un chasquido. La movilización es el movimiento pasivo de las articulaciones para reducir el dolor o recuperar la amplitud.

En TM, la intervención va más allá de la ejecución de una acción específica, siendo más bien un proceso que incluye la valoración, la planificación y la elección de las técnicas a realizar adecuadas para ese paciente en ese momento y la evaluación de su ejecución [18]. Por lo que nos podemos encontrar con diferentes situaciones durante un

tratamiento de TM: I) se pueden utilizar diversas técnicas durante la misma sesión de tratamiento; II) se pueden utilizar diferentes tipos de técnicas para diferentes pacientes que presentan la misma condición patológica. El tratamiento puede diferir entre pacientes debido a la alta variación entre la misma afección (por ejemplo, el dolor lumbar puede tener múltiples etiologías dependiendo de cada paciente, requiriendo así abordajes individuales para cada persona); III) las técnicas pueden variar durante el curso del tratamiento, siendo necesarias una técnicas en una primera sesión y otras en el seguimiento para la adecuada progresión del tratamiento [4]; IV) diferentes disciplinas de TM (fisioterapia, osteopatía, quiropráctica, masoterapia) pueden utilizar diferentes tipos de técnicas de TM en su enfoque terapéutico y combinar la TM con otras intervenciones [1,9]; y V) la regulación educativa de las profesiones de TM varía entre ellas y puede ser diferente entre países [37].

Esto implica que los tratamientos para abordar una determinada condición patológica en la práctica real puedan combinar diversas técnicas diferenciadas para cada paciente, e incluso variar durante el curso de las diversas sesiones de tratamiento. Este escenario dificulta poder plasmar en investigación la realidad de la práctica clínica, y evaluar de forma científica este tipo de intervenciones.

1.2 INVESTIGACIÓN EN TERAPIA MANUAL

1.2.1 Estado de la investigación en Terapia Manual

La producción científica en TM aumentó considerablemente en los últimos años hasta alcanzar un pico en el 2020 [38], decreciendo progresivamente en los últimos 4 años. Usando el término Mesh “Musculoskeletal Manipulations” en una búsqueda bibliográfica en la base PubMed hasta noviembre de 2024, se encontraron un total de 19.294 referencias. De éstas, en los últimos 5 años se publicaron 2709, de las cuales 895 fueron estudios clínicos aleatorizados (ECAs).

1.2.2 El ensayo clínico aleatorizado en Terapia Manual

El ECA es el diseño metodológico por excelencia para demostrar los efectos terapéuticos de una intervención. La mayor parte de su diseño está destinado a minimizar el riesgo de sesgo, es decir, a garantizar la validez interna del estudio (fiabilidad de los resultados) [39,40]. Esto se logra, principalmente, a través de la asignación aleatoria y encubierta de las intervenciones, el enmascaramiento de las mismas y/o la evaluación de forma cegada de los resultados, la elección de una intervención control adecuada y un análisis de los resultados por intención a tratar (ITT) que considere adecuadamente las pérdidas y abandonos [41,42]. Sin embargo, las características metodológicas para garantizar la validez interna pueden comprometer la validez externa (generalización y aplicabilidad de los resultados) del estudio [21,43]. Por otro lado, sacrificar la validez interna para aumentar la validez externa puede resultar en resultados clínicamente relevantes, pero de validez cuestionable. Conseguir un equilibrio entre la validez interna y externa es la principal dificultad de los ECAs, pero especialmente complicado en las ICs y, particularmente, en TM [19–21,39]. Los ECAs altamente controlados y/o con diseños rígidos o muy experimentales entran en conflicto con las características intrínsecas de las ICs ya explicadas anteriormente [19].

1.2.2.1 Estudios de eficacia y de efectividad

Los ECAs pueden tener dos actitudes en función de lo que el estudio pretende investigar y en qué condiciones se pretende realizar el estudio. Los estudios de **eficacia** (también llamados estudios exploratorios o explicativos) se realizan normalmente en condiciones altamente controladas para investigar la relación causal entre una intervención y sus efectos fisiológicos [44]. En cambio, los estudios de **efectividad** (también llamados estudios pragmáticos) tienen la intención de informar sobre las consecuencias de la elección de una intervención sobre otra y se realizan habitualmente en condiciones de la práctica clínica real [44,45]. La **Tabla 2** muestra las principales diferencias entre los estudios de eficacia y los de efectividad.

Tabla 2. Diferencias entre estudios de Eficacia y de Efectividad. (Adaptada de Thorpe et al.2009 y Trewen & Zwarenstein et al. 2009 [46,47])

	EFICACIA	EFFECTIVIDAD
Pregunta de investigación	¿Funciona la intervención?	¿Funciona la intervención en la práctica real?
Validez	Alta validez interna.	Alta validez externa.
Entorno	Con muchos recursos.	Práctica clínica habitual.
Participantes	Criterios estrictos de inclusión y exclusión para reclutar a una población de pacientes definida y homogénea.	Poca o ninguna exclusión más allá de la condición de interés independientemente de riesgo de participación y/o comorbilidades.
	Se suele excluir aquellos sujetos que pueden afectar negativamente al efecto de la intervención o que tendrán poca adherencia.	
Intervención	Se aplica estrictamente y se vigila de cerca su cumplimiento.	Se aplica de forma flexible como se realizará en la práctica habitual, dejando libertad a quien realiza la intervención en cómo aplicarla.
	Profesionales especializados y experimentados con alto grado de éxito en la intervención o a los que se les imparte formación específica	Es aplicada por todos los profesionales y en todos los entornos clínicos clínicos, independientemente

	sobre la intervención.	de sus conocimientos.
Intervención control	Habitualmente placebo/sham control.	Práctica clínica habitual (usual care).
Desenlaces	Normalmente variables fisiológicas.	Habitualmente variables de resultado relevantes para el paciente.

Tanto los ECAs de eficacia como los de efectividad son necesarios para evaluar las intervenciones en salud [48,49]. Sin embargo, en muchas ocasiones, la distinción entre ambos tipos de estudio no es una cuestión simplemente dicotómica. Un mismo estudio puede contener características más relacionadas con la eficacia o con la efectividad, para garantizar algún aspecto metodológico en concreto [45,47]. Lo más importante para distinguir la naturaleza del estudio es conocer la intención del estudio [50], que normalmente ayudará a decidir el enfoque metodológico más adecuado. No obstante, aparte de la superposición metodológica y conceptual, la terminología no siempre se utiliza de forma coherente en la literatura, llevando a malas interpretaciones semánticas entre los términos ‘eficacia’ y ‘efectividad’ [51–53]. Varios autores han remarcado la importancia de esta incongruencia semántica, más allá de la académica, pudiendo tener implicaciones importantes a nivel de la comunicación médica [51]. Interpretar erróneamente la terminología puede afectar a la interpretación de los estudios y el sentido de las guías de práctica clínica, contribuyendo a aumentar la literatura que no es útil y/o relevante en investigación clínica [49,51,52]. Si bien en la investigación farmacológica se han descrito 4 fases específicas para evaluar la seguridad, eficacia, efectividad e implantación de un fármaco, esta clasificación no existe de manera formal para las intervenciones no farmacológicas [54]. Esto puede ser problemático porque la falta de pruebas rigurosas puede llevar a la utilización de intervenciones ineficaces y, por otro

lado, a malinterpretar los términos de eficacia y efectividad.

De los 895 ECAs sobre TM publicados en los últimos 5 años, 28 contenían el término “eficacia” en el título y 77 el de “efectividad”. Este dato ya nos muestra cómo los estudios de efectividad predominan frente a los de eficacia en el campo de la TM o cómo el término Efectividad se usa más frecuentemente en el campo de la TM que el de eficacia. De estos estudios, 7 utilizaron, además, el término pragmático en el título.

Para decidir qué diseño de investigación resulta más adecuado, los investigadores deben evaluar el campo de investigación en su conjunto e identificar qué pregunta de investigación es la más adecuada. ¿Existen todavía dudas sobre la eficacia o, por el contrario, resulta más adecuado llevar a cabo un estudio más pragmático para comprobar cómo se reproducen los efectos en diferentes circunstancias? (por ejemplo, con una menor estandarización de la intervención, una población más amplia, por otros proveedores, etc.). Se han propuesto diferentes estrategias y modelos alternativos al ECA para evaluar las ICs de una forma más precisa [22,55,56]. Entre estas estrategias para mejorar la aplicabilidad de los resultados, los estudios pragmáticos (de efectividad) se han posicionado como un diseño apropiado en el campo de la TM [20,21,57].

1.2.3 Dificultades metodológicas de la investigación en Terapia Manual

A pesar del aumento de la literatura, la investigación en TM está considerada de baja calidad, habiendo sido reportadas numerosas limitaciones y/o debilidades en la literatura científica [1,58–66]. Al realizar investigación en ICs se pueden encontrar retos metodológicos que pueden repercutir en la calidad de los estudios, como la dificultad de aislar sus componentes para realizar estudios de eficacia, identificar sus mecanismos de acción, evaluar la contribución de los efectos no específicos/contextuales y/o comprender su efectividad en la práctica diaria [19,21,22]. Concretamente en la investigación en TM se han descrito ampliamente dificultades debido a las particularidades propias de este tipo de intervención. Entre ellas se puede destacar las

18

siguientes:

1. Intervención experimental:

- a. **Dificultad en protocolizar su diseño:** Diversos autores se han cuestionado si, a parte de la constatación de la baja calidad de la investigación, es posible que las conclusiones poco determinantes de los estudios de TM vengan relacionadas con el hecho de tener que estandarizar en exceso las intervenciones [17,67]. Las intervenciones de TM están centradas en las necesidades específicas de cada paciente, individualizando al máximo el tratamiento a las características de cada persona en un momento dado, sus preferencias, su condición clínica, comorbilidades, etc. Este hecho desafía enormemente la estandarización de la intervención, alejándola de la práctica real [4,13] y dificultando a su vez la replicación del estudio [1,4]. Además, los diferentes tipos de poblaciones, los diferentes contextos del paciente, el lugar o ámbito de ejecución de las intervenciones (hospitales, clínicas privadas, entornos deportivos...) y las distintas disciplinas profesionales que usan la TM (con diferentes niveles de aprendizaje y experiencia) hace que sea difícil establecer un lenguaje estandarizado para describir las intervenciones, dificultando, de nuevo, poder protocolizar y replicar su ejecución [68,69].
- b. **Ejecución de las intervenciones:** La gran variabilidad existente en los parámetros de ejecución de las intervenciones manuales (dosis, cantidad de presión ejercida, velocidad, área donde se contacta, tiempo de ejecución), dificultan la definición y/o descripción objetiva de las mismas. Además, la especificidad de las técnicas (relación entre el lugar donde se aplica la técnica y el resultado clínico obtenido) se ha puesto en duda recientemente [70]. Por otro lado, la experiencia del terapeuta también puede afectar a la efectividad de la técnica [25,26]. Esta gran variabilidad en diferentes aspectos de la ejecución de las intervenciones puede condicionar la efectividad de las mismas [4,71,72], ya que una intervención aplicada por un mismo terapeuta en

diferentes momentos en el tiempo puede variar en términos de fuerza aplicada y duración, por ejemplo. Por este motivo, la comparación entre diferentes terapeutas es difícil de estandarizar. Esta gran variabilidad inter e intra-terapeutas, puede afectar tanto al diseño de la intervención como a su reproducibilidad en otros estudios [71,72].

2. Relación entre la intervención y sus resultados:

Si bien se ha estudiado ampliamente los mecanismos de acción de la TM, no hay una evidencia clara respecto a cuáles son los efectos específicos atribuibles a la TM y su correlación con los resultados clínicos, dificultando la determinación de qué factores explican los resultados obtenidos por la intervención [68]. Al valorar el efecto de la intervención pueden coexistir varios elementos activos que dificulten aislar cuál de los elementos está relacionado con la variable de resultado [55]. Estos aspectos pueden condicionar la incertidumbre clínica, la especificidad de la intervención y la independencia del contexto, que son principios requeridos cuando se quiere realizar un ensayo clínico [55]. Los ECAs pueden demostrar que un grupo de intervención mejora sus resultados en comparación a otro grupo de control, pero no explican el por qué mejoran los resultados [73]. Para ello se necesitaría un análisis de mediación causal y analizar variables intermedias que pueden tener relación entre la intervención y el efecto obtenido. Estas variables intermedias son difíciles de controlar en los pacientes usuarios de la TM debido a su elevada heterogeneidad.

3. Homogeneidad de los grupos:

Una de las características principales de las personas que requieren tratamiento de TM es la heterogeneidad en sus presentaciones clínicas y sus limitaciones funcionales [67,74]. Esto se contrapone a uno de los requerimientos metodológicos de los ECAs que es la homogeneidad de los grupos de estudio que asegure su comparabilidad basal. Limitando los criterios de inclusión se asegura más validez interna del estudio. No

obstante, se aleja de la realidad de los pacientes de la práctica clínica real, donde en este campo en particular, suelen coexistir muchas comorbilidades, diferentes causas que explican la condición patológica que presentan los pacientes y diferentes factores psicosociales [75,76].

4. Intervención control:

Las dificultades para diseñar una intervención control en el campo de la TM afecta a los tres tipos de intervención control posibles:

- a. **Placebo/sham:** las dificultades para describir la intervención experimental suponen un gran reto para poder definir un grupo control placebo o sham que cumpla con los requisitos que debe tener un placebo [18,54]. La intervención control, supuestamente, debería ser igual que la intervención experimental, pero sin su efecto terapéutico. Esto, en investigación farmacológica es fácil de conseguir, pero en las ICs y, especialmente, en la TM donde la intervención comprende la interacción verbal y física entre el paciente y el terapeuta, es un reto que ha despertado gran interés en la comunidad científica [25,54,66,77,78].
- b. **Manejo habitual (“usual care”):** el uso de la TM por varias disciplinas profesionales y contextos en los que se aplica dificulta estandarizar un “usual care” que va a depender del contexto sanitario particular de cada disciplina manual en particular e incluso a cada país.
- c. **Otra intervención activa:** comparar la intervención experimental con otra intervención activa puede estar muy influenciado por las expectativas y/o experiencias previas que los pacientes y/o terapeutas tengan sobre las dos intervenciones que se comparan [18].

5. Tamaño de la muestra:

La literatura muestra que, habitualmente, los estudios de TM se realizan con muestras pequeñas de sujetos [58,79–81]. Esto puede dificultar el equilibrio de los valores o característica basales de los grupos de estudio y su comparabilidad, así como también el poder estadístico para detectar un efecto real de la intervención [58].

6. Enmascaramiento:

El enmascaramiento es un componente fundamental de los ECAs para minimizar el sesgo de medición y obtener estimaciones fiables del efecto del tratamiento. Sin embargo, el enmascaramiento puede ser particularmente desafiante en intervenciones no farmacológicas [41,82,83].

- a. **Enmascaramiento de los pacientes:** El enmascaramiento de los participantes puede ser complicado dependiendo del grupo control que se utilice en el estudio. Cuando se utiliza un grupo control placebo/sham, la dificultad radica en encontrar una intervención convincente que se asemeje al máximo a la intervención experimental, pero sin ningún efecto terapéutico [54]. Al utilizar como intervención control otra intervención activa o el manejo habitual (“usual care”), es también probable que no se consiga el enmascaramiento si no se oculta a los participantes información parcial relativa al estudio [83]. Tanto para un tipo de control u otro, las expectativas de los participantes pueden influir significativamente en los resultados [14,30,83]. Si bien es difícil asegurar un enmascaramiento eficaz, existen estrategias para verificar el éxito del enmascaramiento testándolo al final del estudio [82,84]. No obstante, existe un extenso debate sobre la relevancia del enmascaramiento dependiendo de la fase de desarrollo en que se encuentra la IC, la pregunta de investigación que se quiere responder o el desenlace que se evalúa [83]. Para evaluar el efecto puro de una intervención, el enmascaramiento es esencial [41]. Sin embargo, si el objetivo es evaluar la efectividad en un contexto

real, donde pacientes y profesionales son conscientes de las opciones de tratamiento, un diseño no enmascarado podría ser más apropiado [83].

- b. **Enmascaramiento de los terapeutas:** El enmascaramiento de los terapeutas que imparten la intervención es imposible en el campo de la TM.
- c. **Enmascaramiento del/la evaluador/a externo:** Es altamente recomendado que un evaluador externo evalúe los resultados de forma cegada cuando el enmascaramiento de los participantes y terapeutas incluidos en el estudio no es posible [55]. No obstante, cuando la variable de resultado es medida a través de una PROM, el mismo paciente se convierte en evaluador, con lo que, en estos casos, si el enmascaramiento del paciente no se puede conseguir, este aspecto metodológico puede quedar comprometido también.

7. Variables de Resultado:

Un aspecto importante de la medición de resultados en TM es el hecho de medir resultados a largo plazo [18]. Si bien se pueden medir efectos a corto plazo, lo interesante en estas ICs que pueden afectar a varias dimensiones, es ver cómo estos efectos a corto plazo se mantienen en el tiempo, e incluso son parte de cambios en el proceso de salud y los hábitos de los pacientes [18]. Medir a largo plazo puede comportar costes y una logística complicada difíciles de asumir por la investigación en el campo de la TM.

8. Generalización de los resultados:

La validez externa se refiere al grado en que los resultados del estudio pueden ser aplicados más allá de la muestra de población estudiada, a otras personas distintas y/o en diferentes localizaciones o ámbitos. La limitación más conocida del ECA es la dudosa transferibilidad de sus resultados. Esta brecha entre investigación y clínica es la principal crítica que recibe la práctica basada en la evidencia por parte de los clínicos [21,85],

sobre todo en el campo de las ICs [43].

Todos los elementos que garantizan la validez interna de los estudios suelen afectar la validez externa del mismo. Estos elementos, como las características de los pacientes incluidos, la intervención realizada y las variables medidas, si se alejan de la práctica habitual, pueden condicionar la transferibilidad de los resultados del estudio a la práctica real. Precisamente, una revisión sistemática publicada recientemente ha demostrado que los resultados de los estudios de TM son poco generalizables y aplicables a diferentes poblaciones y escenarios [48].

9. Financiación:

La TM es parte de la atención asistencial de diferentes profesiones (fisioterapia, osteopatía, quiropráctica, acupresión), no obstante, estas profesiones no están todas incluidas en los sistemas de salud pública. Este hecho implica que muchos de los estudios que se realizan sobre TM no se enmarcan dentro de un entorno sanitario público con financiación por parte de organismos oficiales, siendo los propios investigadores los que muchas veces acarrean con los gastos que implica realizar investigación o consiguiendo recursos muy limitados [86,87]. Esta falta de recursos puede afectar directamente al tamaño de la muestra que se necesita ya la duración del estudio [88].

1.2.4 Propuestas metodológicas adaptadas a la Terapia Manual

No todas las preguntas de investigación se pueden contestar adecuadamente por medio de un ECA [73]. Dependiendo de cuál sea la pregunta, otros diseños metodológicos pueden resultar más indicados. Estudios de caso-control, de cohortes prospectivas, revisiones sistemáticas y meta-análisis y estudios observacionales, entre otros, pueden ser diseños con alta capacidad de responder a preguntas de investigación de forma más adecuada que un ECA [89].

Si bien se han realizado muchos progresos en cuanto al desarrollo de guías o directrices para ayudar a los investigadores a mejorar la evaluación de las ICs, aún sigue habiendo abundante literatura científica que no es de mucha utilidad para ayudar a tomar decisiones clínicas. Es muy diferente evaluar la efectividad de un fármaco que evaluar la efectividad de una IC y es por ello que, a veces, un ensayo clínico aleatorizado (ECA) puede ser un diseño no fácil de implementar e incluso, a veces, impracticable [90]. Diversos autores defienden el uso del ECA para evaluar ICs [17,54], mientras que otros han propuesto diferentes modelos que permitan evaluar las ICs de una forma más pertinente más allá del ECA individual ya que, al ser contexto-dependientes, su transferibilidad puede no ser evidente [16,55,56].

El Medical Research Council junto con el National Institute for Health Research elaboraron un marco conceptual para evaluar las ICs donde se prioriza la idea de que la efectividad de la intervención es la clave para su evaluación [22] (Figura 4). No obstante, con el tiempo, se plantearon evaluar otros aspectos además de la efectividad, ya que centrarse solo en la efectividad de la intervención única y exclusivamente, podría conllevar a intervenciones no aplicables y/o no útiles en la práctica real. Para lograrlo, la investigación debía involucrar desde el principio a pacientes, médicos y diferentes partes implicadas. Esto implicó pasar de preguntar únicamente si una intervención es efectiva, a analizar también si será aceptada, aplicada, rentable, escalable y adaptable a distintos contextos, y a cómo lograrlo [16].

Para ello, identificaron cuatro perspectivas para orientar el diseño y la realización de investigaciones sobre ICs: I) perspectiva de eficacia (determinar hasta qué punto la intervención produce los resultados deseados en entornos experimentales o ideales), II) perspectiva de efectividad (determinar hasta qué punto la intervención produce los resultados deseados en entornos del mundo real), III) perspectiva basada en la teoría (comprender cómo se produce el cambio, incluida la interacción de los mecanismos y el contexto) y IV) perspectiva de sistemas (esta perspectiva considera la intervención como una parte de un sistema complejo y busca comprender cómo el sistema y la intervención

se relacionan entre sí) [16]. Cada perspectiva puede responder a una pregunta de investigación en concreto, o que varias se solapen y puedan contestar a una misma pregunta de investigación. Este marco conceptual (**Figura 4**) describe cuatro fases de investigación en las ICs: desarrollo o identificación de la intervención, viabilidad, evaluación y aplicación. Estas fases tienen unos elementos comunes:

- Considerar el contexto: Los efectos de una intervención compleja suelen depender, en gran medida, del contexto, de manera que una intervención eficaz en algunos entornos podría ser ineficaz o incluso perjudicial en otros.
- Desarrollar, refinar y probar la teoría: Describe cómo se espera que una intervención produzca sus efectos y en qué condiciones.
- Involucrar a las partes interesadas en la intervención: Incluyen a las personas a las que se dirige la intervención, las que participan en su desarrollo o ejecución, o aquellas cuyos intereses personales o profesionales se ven afectados.
- Identificar las principales incertidumbres: En cada fase del proceso de investigación se podrían responder a muchas preguntas.
- Perfeccionamiento de la intervención: Dentro de cada fase de la investigación de la intervención compleja, y en la transición de una fase a otra, la intervención podría necesitar ser refinada.
- Consideraciones económicas: La evaluación económica debe ser un componente básico de todas las fases de la investigación de la IC.

Figura 4. Marco conceptual para evaluar las Intervenciones complejas

MARCO CONCEPTUAL PARA EVALUAR LAS INTERVENCIONES COMPLEJAS		
PERSPECTIVAS	FASES	ELEMENTOS COMUNES
1- Eficacia 2- Efectividad 3- Basada en la teoría 4- De sistemas	1- Desarrollo 2- Viabilidad 3- Evaluación 4- Aplicación	1- Considerar el contexto 2- Desarrollar, refinar y probar la teoría 3- Involucrar a las partes 4- interesadas en la intervención 5- Identificar las principales incertidumbres 6- Perfeccionamiento de la intervención 7- Consideraciones económicas

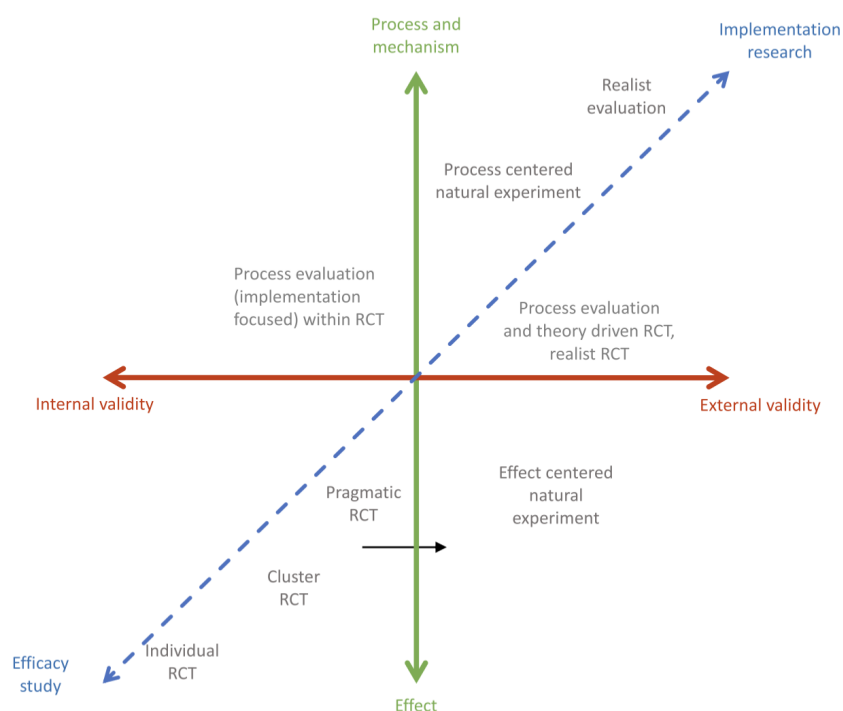
En cada fase se pueden desarrollar estudios con cualquier perspectiva teniendo en cuenta los elementos comunes y pudiendo usar diferentes diseños metodológicos. Más allá del ECA convencional (de grupos paralelos), se podría considerar que los diseños adaptativos, los ensayos SMART (ensayos aleatorizados aleatorios secuenciales), los ensayos n-de-1 y los diseños híbridos de eficacia e implementación podrían ser útiles para desarrollar métodos que ayuden a valorar la investigación sobre la eficacia de las ICs. Un enfoque puramente cuantitativo, utilizando un diseño experimental sin elementos adicionales, como una evaluación de los procesos, rara vez es adecuado para la investigación de ICs. En cambio, enfoques cualitativos y mixtos pueden ser necesarios para responder preguntas más allá de la efectividad.

Este marco cuestiona la idea de que la efectividad sea siempre el objetivo fundamental de la evaluación. Para algunas preguntas sobre ICs, investigar sobre su eficacia o efectividad puede ser el enfoque óptimo y un ECA proporcionará el mejor diseño para lograr una estimación imparcial, mientras que para otras preguntas, perspectivas y diseños alternativos pueden ayudar a responderlas mejor.

En esta línea, Minary et al., en su revisión de estudios que evalúan ICs, concluyó que no

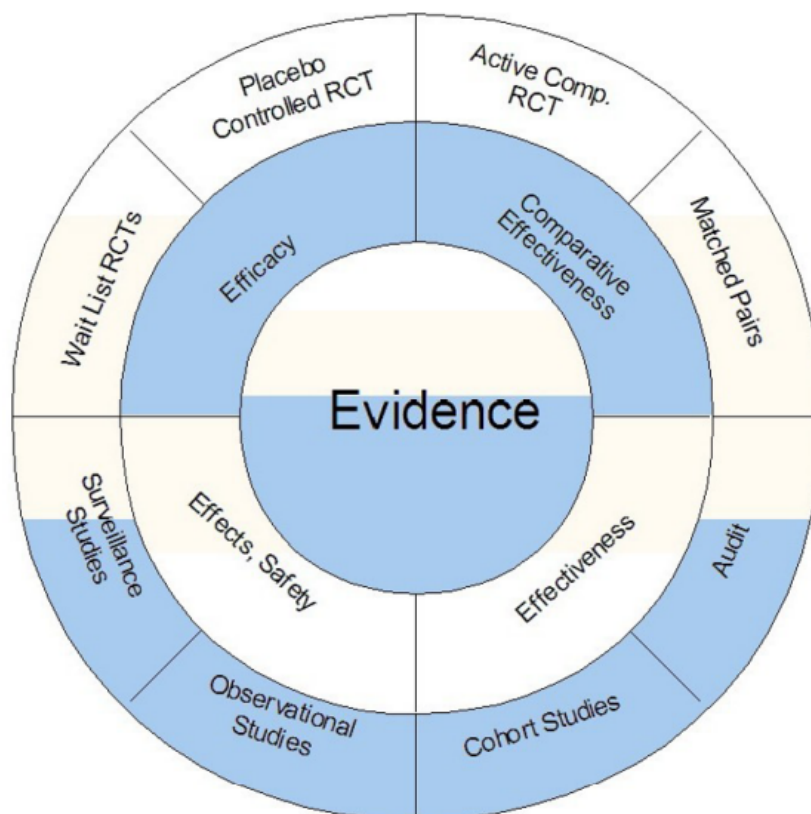
hay un método ideal para evaluar las ICs y propuso adaptar el diseño del estudio en función del tipo de investigación que se quiere realizar (**figura 5**) [56].

Figura 5. Modelo propuesto por Minary et al. para evaluar las ICs [56]



En la misma línea, Walach et al. plantearon otro marco conceptual con un enfoque circular donde describen las dificultades que las ICs encuentran para cumplir con las premisas necesarias cuando se pretende realizar un ECA. En la **Figura 6** se muestra el modelo de Wallach et al. mostrando los diferentes diseños metodológicos en función de la dimensión de la intervención que se quiere estudiar [55].

Figura 6. Modelo para evaluar las ICs de Wallach et al. [55]



Por otro lado, diversos autores han identificado las dificultades metodológicas al realizar investigación de una IC y propuesto estrategias para ayudar a que su impacto repercuta lo mínimo en la calidad de los estudios. Por ejemplo, entre la comunidad científica que investiga sobre el dolor, se considera que los estudios de efectividad comparativa pueden ser un buen método para estudiar las ICs siempre que existan estudios de eficacia controlados con una intervención sham [91]. Otro ejemplo es el que Gerard Álvarez propuso en su tesis doctoral, donde planteó una serie de recomendaciones para paliar las dificultades metodológicas de los estudios de eficacia en TM y contribuir a mejorar la calidad de los estudios en este campo [92]. Esta propuesta, junto a otros recursos publicados recientemente [54,93], están recogidas en la **tabla 3**.

Tabla 3. Recomendaciones para paliar las dificultades metodológicas de los estudios de eficacia en Terapia Manual [92]

Tamaño de la muestra	<ul style="list-style-type: none"> • Incluir y mejorar el reporte del cálculo de la muestra. • Especificar el valor Mínimo de Diferencia Clínica. • Asegurar métodos robustos y mejorar el reporte cuando las muestras sean pequeñas.
Ausencia de placebo validado	<ul style="list-style-type: none"> • Mejorar las estrategias placebo y su procedimiento. • Mejorar el reporte de estas intervenciones. • Uso de la guía TIDIER placebo. • CoPPS statement: recomendaciones para el desarrollo, implementación y reporte de las intervenciones control en estudios de eficacia de intervenciones no farmacológicas.
Enmascaramiento	<ul style="list-style-type: none"> • Asegurar enmascaramiento simple cegando los evaluadores. • Obtener una credibilidad del efecto del tratamiento • Valorar las preferencias y expectativas de los participantes. • Enmascarar los participantes sobre la hipótesis del estudio. • Diseñar e implementar medidas

	<p>para evaluar el éxito del placebo.</p> <ul style="list-style-type: none"> • Asegurar reportar las estrategias usadas para minimizar el 'sesgo de detección cuando el enmascaramiento no es posible.
Duración del estudio y seguimiento	<ul style="list-style-type: none"> • Reportar de forma adecuada las pérdidas y sus razones. • Priorizar el análisis por Intención a tratar (ITT). • Reportar de forma adecuada el plan de análisis y como se tratarán los datos incompletos.
Intervención experimental	<ul style="list-style-type: none"> • En la evaluación de las técnicas aplicar protocolos. • En la evaluación de los tratamientos: establecer marcos conceptuales, algoritmos o diagramas de trabajo. • Usar PRECIS-2 en la fase de protocolo. • Usar TIDIER para la descripción de la intervención.
Variables de resultado	<ul style="list-style-type: none"> • Usar la extensión PRO Consort para el reporte de "patient reported outcomes" en ECAs. • Reportar la variable primaria adecuadamente. • Registrar el protocolo en registros públicos. • Reportar el efecto inter-grupos

	<p>aparte de intra-grupos.</p> <ul style="list-style-type: none"> • Incluir intervalos de confianza. • Usar paquetes de variables protocolizadas (iniciativa COMET). • Usar PROMs. • Asegurar un buen reporte gráfico de los resultados.
--	--

1.3 ENSAYOS CLÍNICOS ALEATORIZADOS PRAGMÁTICOS

Los ECAs pragmáticos (o de efectividad) (EP) son aquellos estudios diseñados para determinar los efectos de una intervención en las condiciones donde habitualmente se aplican (práctica real), con el objetivo de maximizar la aplicabilidad y generalización del efecto de la intervención. Esta distinción entre los estudios pragmáticos y los de eficacia fue propuesta por primera vez por dos autores franceses, Schwartz y Lellouch, en un artículo publicado en 1967 [44]. En él señalaron la limitada aplicabilidad de algunos estudios.

En los EPs la intervención se compara con otras intervenciones realizadas en el propio escenario de la práctica habitual. Miden un amplio espectro de desenlaces principalmente centrados en el paciente y suelen ser muy heterogéneos en cuanto a las características de los pacientes, el escenario clínico y el modo cómo se aplica el tratamiento. En contrapartida, deben tener una muestra grande y deben ser simples en su diseño [41,47].

Una de las principales críticas a la medicina basada en la evidencia desde el punto de vista clínico es la "baja capacidad de la evidencia científica para parecerse a entornos clínicos reales" [21,85]. Esta carencia de relevancia es la razón principal por la que, a

veces, los profesionales no adoptan intervenciones basadas en la investigación [43] y podría explicar por qué los tratamientos que han demostrado ser beneficiosos en estudios son infrutilizados en la práctica clínica habitual [94,95]. Además, más del 85% de los esfuerzos en investigación biomédica se desperdician debido a la poca transferencia de los resultados de la investigación a la práctica [96]. Esta falta de atención a la validez externa está aumentando la brecha entre la cantidad de investigación disponible y la poca relevancia que tiene para los profesionales [43]. Las revistas científicas, las agencias reguladoras y financieras, los comités éticos, la industria farmacológica y las regulaciones gubernamentales parecen no dar prioridad a la validez externa de los estudios [43,47,94]. Es por ello que varios autores han mostrado interés en este tipo de estudio e incluso se han desarrollado herramientas para ayudar a su diseño e implementación [21,47,50,95,97,98].

1.3.1 Diseño de ensayos clínicos aleatorizados pragmáticos

Thorpe et al. crearon una herramienta llamada "PRagmatic-Explanatory Continuum Indicator Summary" (PRECIS) en 2008 para ayudar al diseño de ECAs con una actitud pragmática [46]. Fue revisada y mejorada en 2015 por un amplio grupo de investigadores a partir de los comentarios recibidos, creando la herramienta PRECIS-2 [95]. El objetivo era ayudar a los investigadores a diseñar estudios según los objetivos del estudio. PRECIS-2 consta de 9 dominios: elegibilidad, reclutamiento, configuración, organización, flexibilidad (intervención), flexibilidad (adherencia), seguimiento, resultado primario y análisis primario (**Tabla 4**). Cada dominio se puntúa en una escala tipo Likert de 5 puntos, con una puntuación posible desde "1" -que se considera muy explicativa- hasta "5" -que se considera muy pragmática (práctica habitual)-. Las puntuaciones de cada dominio se representan en un gráfico radial llamado rueda PRECIS-2 (**Figura 7a**). Los ensayos con puntuaciones más explicativas muestran una rueda más estrecha hacia el centro y los ensayos que contienen puntuaciones más pragmáticas tienen una rueda más ancha hacia la periferia, como se puede observar en la **Figura 7b**. Las puntuaciones para cada dominio pueden variar a lo largo del continuo explicativo-pragmático. Además, en el

33

propio ECA pueden coexistir dominios puntuados como pragmáticos con otros puntuados como explicativos [95]. Actualmente, está en fase de desarrollo la versión PRECIS-3 de la herramienta [99]. Esta nueva versión se está desarrollando a partir de los comentarios y sugerencias publicados en la literatura sobre la herramienta PRECIS-2 [99]. Se valorará la inclusión de un dominio para el grupo control, un dominio para el enmascaramiento, así como mejoras en la descripción de los dominios ya existentes [99]. Varios estudios han utilizado la herramienta PRECIS-2 para evaluar la actitud pragmática, ya sea en la fase de diseño o de ejecución de los estudios, concluyendo que es una herramienta válida para este fin [100–107].

Tabla 4. Dominios de la herramienta PRECIS-2 [95]

DOMINIO	EXPLICACIÓN
1.ELEGIBILIDAD	Cómo se parecen los participantes del estudio a los que recibirían la intervención en la realidad, tanto los participantes como el sistema sanitario.
2.RECLUTAMIENTO	Cuánto esfuerzo extra debe realizarse para reclutar participantes en relación al que se haría en la práctica habitual.
3.ENTORNO	Cómo de diferente es el escenario dónde se realizará el estudio del escenario al usado habitualmente
4.ORGANIZACIÓN	Cómo de diferentes son los recursos utilizados, la experiencia del terapeuta y la organización de cómo se realizará la intervención respecto a la práctica habitual.
5.APLICACIÓN DE LA INTERVENCIÓN	Cómo de diferente es la flexibilidad con la que se lleva a cabo la intervención comparada con la práctica habitual.
6.ADHERENCIA A LA INTERVENCIÓN	Cómo de diferente es la flexibilidad en cómo los participantes se adhieren a la intervención respecto a la práctica habitual.
7.SEGUIMIENTO	Cómo de diferente es cómo se mide y cómo se siguen los participantes respecto a la práctica habitual.
8.VARIABLES DE	Hasta qué punto el resultado principal es relevante para los

RESULTADO	participantes.
9. ANÁLISIS ESTADÍSTICO	Hasta qué punto se han incluido todos los datos en el análisis estadístico de los resultados.

Figura 7a. Rueda PreciS-2 (de www.precis2.org).

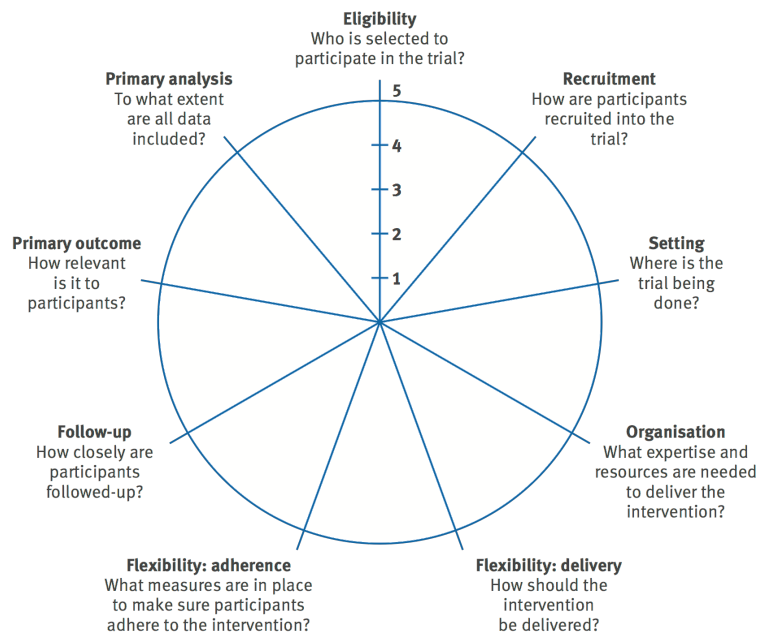
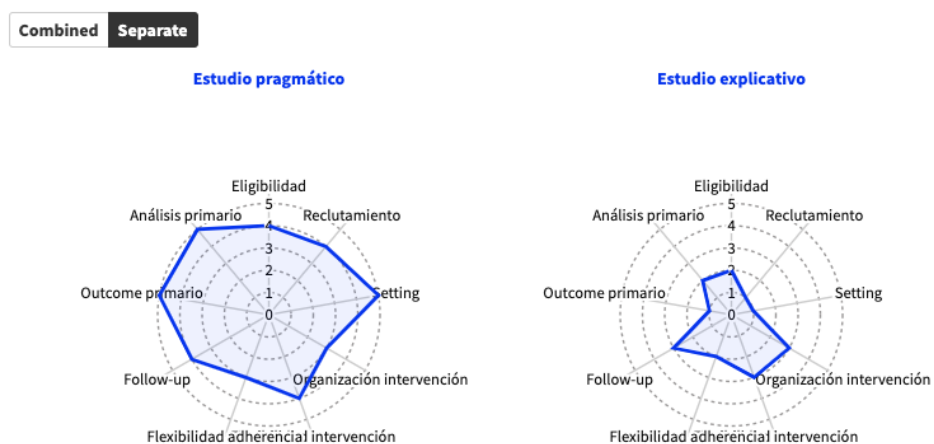


Figura 7b. Rueda PreciS-2 Estudio pragmático vs estudio explicativo



1.3.2 Evaluación retrospectiva de la actitud pragmática de los ECAs

Aunque PRECIS-2 se creó para ser utilizada de forma prospectiva durante la fase de diseño, los investigadores han utilizado esta herramienta en la literatura para evaluar de forma retrospectiva las características pragmáticas de algunos diseños [32,50,98,100–105,107–110].

Recientemente, se ha señalado que, cuando se evalúa la actitud pragmática de un estudio, se debe poner especial atención en detectar su intención más allá de obtener una puntuación cuantitativa de su pragmatismo [50,111]. Al conocer la intención del estudio, cabe esperar que algunos dominios sean más pragmáticos que otros. De hecho, algunos autores afirman que la actitud pragmática es específica de cada dominio y que las calificaciones PRECIS-2 requieren contexto para evaluarse debidamente [32]. De hecho, uno de los principios fundamentales de la validez externa es que el contexto importa [43,112]. El contexto ayudará a entender mejor la actitud del estudio, evitando la dicotomización entre pragmático y explicativo [50]. Además, el uso de la extensión CONSORT para ensayos pragmáticos [97] es muy recomendable para mejorar como se reporta la información del estudio, contextualizar los métodos y resultados y entender la intención del estudio [50,111]. Esta extensión incluye información específica sobre la relevancia de la intervención y los resultados que el estudio pretende abordar.

1.3.3 Limitaciones de los ECAs pragmáticos

Es necesario matizar que sería un error asumir que debido a que un estudio sea pragmático necesariamente tiene mayor validez externa que un estudio de eficacia. Ford et al. publicaron en 2016 un artículo en el New England Journal of Medicine donde se exponen las fortalezas y limitaciones de los estudios pragmáticos (EPs) así como retos para su implementación [21].

La principal preocupación respecto a los EPs es que sus resultados pueden estar

fuertemente sesgados, ya que al tener un menor control sobre las variables de confusión no es posible determinar con precisión la efectividad real de la intervención [41,113,114]. En segundo lugar, al incluir muestras más heterogéneas para reflejar mejor la práctica clínica habitual, los EPs pueden arrojar resultados engañosos si la intervención resulta ser efectiva solo en un subgrupo específico de participantes [41,114,115]. Una tercera limitación es de tipo logístico y económico: debido al mayor riesgo de sesgo y a la menor sensibilidad, estos estudios suelen requerir tamaños muestrales más grandes que los ensayos explicativos, lo que conlleva mayores costos y una duración más prolongada. Además, Rowbotham et al. han señalado otras limitaciones específicas de los EPs en el contexto del estudio del dolor crónico [115]. Por ejemplo, aunque un diseño pragmático idealmente debería ser multicéntrico, esto puede introducir una gran variabilidad entre centros, lo que dificulta la interpretación de los resultados. Asimismo, aunque se busca no controlar la experiencia de los terapeutas para asemejar los ensayos a la práctica clínica real, en el caso del dolor crónico la notable variabilidad en el conocimiento y las habilidades clínicas puede influir significativamente en los resultados, constituyendo otra fuente de confusión.

Estas preocupaciones, si bien se deben tener en cuenta en todo tipo de intervenciones, son especialmente importantes en las intervenciones no reguladas [98], ya que, a falta de estudios de eficacia de buena calidad, sus resultados pueden ser fuertemente sesgados al tener limitaciones en su validez interna [21,41,113–115]

Por otro lado, existe un amplio debate sobre si algunas características metodológicas específicas de este tipo de estudios son adecuadas para mostrar la práctica habitual [50,99,111,116–119]. Por ejemplo, ¿puede un estudio informar la toma de decisiones clínicas en la atención habitual utilizando una intervención control de placebo o realizándose en un entorno unicéntrico? ¿Las intervenciones estandarizadas son aplicables a un entorno asistencial donde no se aplican protocolos? De hecho, algunos autores se han cuestionado si los estudios publicados en la literatura bajo la etiqueta de "efectividad" tienen suficiente actitud pragmática para parecerse a la práctica clínica e

informar para tomar decisiones en un entorno clínico real [120,121].

1.4 JUSTIFICACIÓN DE LA TESIS

La Terapia Manual (TM) se considera una intervención compleja (IC). Como tal, presenta una serie de dificultades metodológicas intrínsecas que pueden comprometer seriamente el diseño, la ejecución y la interpretación de los estudios de investigación. Estas dificultades han sido objeto de análisis en múltiples trabajos centrados en los estudios de eficacia, y reflejan los desafíos comunes al investigar intervenciones donde intervienen múltiples componentes, alta variabilidad clínica y una considerable dependencia del contexto y del profesional. En general, los estudios sobre TM han mostrado una calidad metodológica limitada, caracterizada por un reporte deficiente e incompleto, lo cual compromete no solo su interpretación, sino también la evaluación rigurosa de su calidad y su aplicabilidad clínica [1].

En respuesta a estas limitaciones, se han propuesto diversos modelos para evaluar de forma más adecuada las ICs, y específicamente en el campo de la TM, se han sugerido estrategias y alternativas metodológicas orientadas a mejorar la calidad del diseño y la evaluación de las intervenciones. Entre estas, los estudios pragmáticos han emergido como una posible vía para abordar la escasa aplicabilidad clínica que caracteriza gran parte de la literatura existente. A diferencia de los ensayos explicativos, los estudios pragmáticos buscan aproximarse a las condiciones del mundo real, aumentando así la validez externa de los resultados. Sin embargo, aunque se ha analizado el grado de pragmatismo en estudios de otras disciplinas médicas, actualmente no existen estudios que evalúen sistemáticamente el pragmatismo en el ámbito de la TM.

Esta tesis se presenta como una continuación y complemento de la tesis doctoral realizada por Gerard Álvarez, titulada *Retos metodológicos en la evaluación de la eficacia de la Terapia Manual* [92]. Ambas investigaciones surgen en respuesta a dos preguntas fundamentales que representan un desafío persistente en el campo de la TM: por un lado, analizar la calidad de la evidencia disponible; por otro, valorar en qué medida esa evidencia es aplicable a la práctica clínica. En su trabajo, Álvarez examinó una muestra

de estudios de eficacia en TM y concluyó que, a pesar del aumento en la producción científica, la calidad sigue siendo deficiente y el reporte metodológico es pobre, lo cual obstaculiza seriamente la evaluación de los resultados y su aplicación clínica. A partir de estas conclusiones, propuso diversas soluciones a los obstáculos metodológicos, incluyendo la necesidad de aumentar el rigor en los estudios de eficacia, considerar variantes metodológicas del ensayo clínico aleatorizado (ECA) y, en especial, explorar alternativas como los estudios pragmáticos.

Sobre esta base, la presente tesis se centra en profundizar en los estudios de efectividad, analizando cómo se conceptualiza y aplica el pragmatismo en el contexto de la investigación en TM. Esta línea de trabajo responde a una necesidad no resuelta en el campo: la de examinar hasta qué punto los estudios de efectividad autodenominados "pragmáticos" cumplen realmente con las características propias de este diseño, y en qué medida los estudios que no se etiquetan como tales podrían también aproximarse al modelo pragmático. Esta reflexión no es meramente semántica, sino que tiene profundas implicaciones metodológicas, clínicas y epistemológicas.

Desde una perspectiva metodológica, evaluar el grado de pragmatismo mediante herramientas específicas como PRECIS-2, junto con el análisis del reporte (a través de la guía CONSORT) y del riesgo de sesgo, permitirá determinar si el uso del término "efectividad" en el ámbito de la TM está respaldado por un diseño coherente y robusto. Desde el punto de vista científico, esta tesis aborda una laguna importante al aportar un análisis sistemático de la calidad metodológica y la aplicabilidad de los estudios de efectividad en TM, contribuyendo así a mejorar la transparencia y el rigor en la investigación. Finalmente, desde una perspectiva práctica, los hallazgos de esta tesis podrán tener un impacto en la forma en que se diseñan, evalúan e interpretan los estudios clínicos en TM, ayudando a que los resultados de la investigación sean más útiles y relevantes para la toma de decisiones en el entorno clínico.

En suma, esta tesis busca no solo analizar el estado actual de la investigación en TM

desde la óptica del pragmatismo, sino también generar propuestas concretas para mejorar el diseño de futuros estudios de intervención en este campo, fortaleciendo la conexión entre la investigación científica y la práctica clínica real.

2. OBJETIVOS

2.1 Objetivo principal

El objetivo de esta tesis es analizar en profundidad las características metodológicas, el grado de pragmatismo, el reporte y el riesgo de sesgo de los ensayos clínicos aleatorizados de efectividad realizados en el campo de la Terapia Manual (TM), con el fin de valorar cuán pragmáticos son realmente los estudios que se autodenominan pragmáticos y evaluar el uso del término “efectividad” dentro de la literatura científica en este ámbito.

2.2 Objetivos secundarios

1. **Describir y analizar el grado de pragmatismo** de los ensayos clínicos aleatorizados (ECAs) de efectividad en Terapia Manual, tanto si se autodenominan pragmáticos como si no.
2. **Evaluar la calidad del reporte metodológico** de los ECAs de efectividad en Terapia Manual, independientemente de su autodenominación como pragmáticos.
3. **Determinar el riesgo de sesgo** presente en los ECAs de efectividad en Terapia Manual.
4. **Valorar críticamente el uso del término “efectividad”** en la literatura científica sobre Terapia Manual, determinando si los estudios que lo emplean cumplen los criterios metodológicos propios de los diseños de efectividad/pragmáticos.

5. **Analizar comparativamente los estudios autodenominados pragmáticos y los no autodenominados**, en relación con su diseño, pragmatismo, reporte y riesgo de sesgo, para identificar posibles inconsistencias, tendencias o malentendidos metodológicos en el uso del enfoque pragmático en Terapia Manual.
6. **Proponer recomendaciones metodológicas** que contribuyan a mejorar el diseño, la evaluación y la aplicabilidad de las intervenciones complejas, en particular la Terapia Manual, dentro de la investigación clínica.

3. MÉTODOS

Con el fin de alcanzar los objetivos de esta tesis, se desarrollaron tres investigaciones cuyos métodos se describen en las publicaciones correspondientes. Estos trabajos sirven de base para estructurar la tesis bajo el formato de compendio de artículos.

Se llevó a cabo una revisión metodológica exhaustiva de ensayos clínicos aleatorizados centrados en evaluar la efectividad de las intervenciones de TM. Esta revisión siguió las directrices del Cochrane Handbook for systematic reviews of interventions [122] y las tres publicaciones derivadas de ella han sido reportados según la guía PRISMA [123]. El protocolo de esta revisión se registró en Open Science Framework (DOI 10.17605/OSF.IO/WKEPZ) en junio del 2020.

A continuación, se presenta un resumen de los métodos de cada uno de los estudios.

3.1 Métodos del primer trabajo

El objetivo principal del primer trabajo consistió en analizar las características y el grado de pragmatismo de los estudios de efectividad autodenominados pragmáticos (como diseño del estudio) en su título y/o resumen en el campo de la terapia manual.

3.1.1 Criterios de elegibilidad de los estudios incluidos

Se incluyeron ECAs publicados que en su título y/o resumen contuvieran los términos "efectividad" (relacionados con una intervención), "pragmático" (relacionados con el diseño metodológico) o "naturalista" (diseño metodológico) para poder identificar aquellos estudios con un diseño metodológico que pretendiera asemejarse al máximo a la práctica clínica habitual [98,124]. Las referencias elegibles debían incluir una técnica

manual o una combinación de técnicas manuales (técnicas de tejidos blandos, movilización o manipulación de articulaciones, masaje, liberación miofascial, manipulación de nervios, tensión/contratensión o acupresión). No hubo restricciones en términos de la población de estudio, los grupos de comparación o las medidas de resultado analizadas. Los criterios de exclusión incluyeron intervenciones experimentales que utilizaban herramientas, dispositivos (electroterapia, kinesiotaping, punción seca, acupuntura), fármacos, ejercicios activos o una combinación de terapias sin la intervención de una TM. También se excluyeron artículos de prensa, protocolos y presentaciones de pósteres/conferencias y estudios que no estuvieran escritos en inglés.

3.1.2 Estrategia de búsqueda

Se realizó una búsqueda exhaustiva en MEDLINE y el Registro Cochrane Central de estudios Controlados (CENTRAL) desde el inicio hasta junio de 2024, utilizando una estrategia de búsqueda que combinó vocabulario controlado con términos MeSH relevantes en el campo de la TM y el diseño de interés (ANEXO 1). La estrategia de búsqueda fue desarrollada por un metodólogo experto que formó parte de nuestro equipo de investigación.

3.1.3 Selección de los estudios

Los registros identificados por las búsquedas se cargaron en el software Rayyan (www.rayyan.ai) [125]. Tras eliminar los artículos duplicados, las referencias fueron revisadas en relación a los criterios de inclusión por dos revisores independientes a partir del título y el resumen, resolviendo los desacuerdos mediante discusión.

3.1.4 Proceso de recopilación y extracción de datos

El grupo revisor estuvo compuesto por 7 revisores. La investigadora principal se emparejó con todos los miembros del grupo revisor para recopilar y analizar los datos de los estudios incluidos por pares. Se resolvieron los desacuerdos mediante consenso.

Previo al proceso de revisión, el equipo realizó un taller formativo con la Dra. Kirsty Loudon, autora de la herramienta PRECIS-2 [95] para mejorar el dominio de la herramienta. Se examinaron, compararon y discutieron tres artículos según los criterios de la Dra. Loudon.

Tres autores diseñaron un formulario de extracción de datos (**ANEXO 2**) que contenía:

1. Elementos bibliométricos de identificación y características específicas del estudio.
2. La intención del estudio dada por los autores: Evaluada mediante las siguientes tres categorías cualitativas preestablecidas por el equipo de revisores y dirigidas a cómo los autores informaron si su objetivo era I) examinar la efectividad de una intervención (sólo informar de cómo responden los participantes al tratamiento, examinar y estimar efectos/cambios); II) cómo se parecía la intervención a la práctica clínica en cuanto a las poblaciones incluidas, el entorno o la administración de la intervención; y III) examinar la efectividad de una intervención realizada en un entorno lo más parecido posible a la práctica clínica real [50,109].
3. La justificación de la intervención dada por los autores: Esta información se buscó en la sección de introducción de cada estudio. Se refiere a la existencia de cualquier información sobre la eficacia de la intervención. Los revisores preestablecieron tres categorías: I) la existencia de experimentos mecanicistas que explicaran los mecanismos de acción de la intervención; II) la existencia de estudios controlados simulados que probaran la intervención frente a un placebo; y III) datos de efectividad comparativa sobre la intervención [126].
4. Intervenciones experimentales y de control: El equipo de investigación utilizó un

proceso de categorización para distinguir entre las intervenciones experimentales y las de control. Esto se hizo para evaluar en qué medida la intervención difería de la atención habitual. Como intervención centrada en el paciente, un enfoque pragmático de la TM requiere una adaptación individual para cada persona y también la adaptación de diferentes técnicas a medida que progresa la enfermedad [4]. Parte del equipo de investigación eran expertos en TM y establecieron una clasificación de la intervención según su capacidad para asemejarse a la práctica clínica. La **tabla 5** hace referencia a las categorías preestablecidas para clasificar las intervenciones experimentales y de control. Las intervenciones experimentales se dividieron entre las que estaban protocolizadas debido a los requisitos del estudio y las elegidas por el terapeuta en función de las necesidades del paciente. Se distinguió entre las intervenciones que utilizaban una sola técnica como intervención experimental, las que utilizaban una combinación de técnicas (p.e técnicas de tejidos blandos, manipulaciones y/o movilizaciones pasivas), o las que combinaban diferentes terapias dentro de la misma intervención (p.e manipulaciones vertebrales y ejercicios vs ejercicios solo)

Tabla 5. Categorías preestablecidas para clasificar las intervenciones experimental y control

CATEGORÍAS DE LA INTERVENCIÓN EXPERIMENTAL		
Combinación de técnicas no protocolizadas	de no	Uso de una combinación de técnicas de TM ya utilizadas en la práctica clínica, pero elegidas por el profesional en función de las necesidades del participante.
Técnica aislada no protocolizada	de no	El profesional elige una única técnica de TM en función de las necesidades del participante.
Protocolo de una combinación de técnicas	de de	Uso de una combinación de técnicas de TM estandarizadas por el protocolo del estudio.
Combinación de terapias no	de no	Uso de una combinación de distintas modalidades de tratamiento en función de las necesidades de los pacientes,

protocolizadas	por ejemplo: TM combinada con ejercicios.
Técnica aislada protocolizada	Uso de una única técnica de TM estandarizada por el protocolo del estudio.
Protocolo de una combinación de terapias	Uso de una combinación de terapias estandarizada para todos los pacientes por el protocolo del estudio.
CATEGORÍAS DE LA INTERVENCIÓN DE CONTROL	
Intervención activa	Uso de otra intervención aceptada o establecida que no se considera más allá de la etiqueta «usual care».
Placebo/sham	Uso de una intervención supuestamente inactiva que se parece al tratamiento que se está probando
Usal care	La intervención que se espera recibir como parte de la atención clínica normal
No intervención	Grupo que no recibe ninguna intervención

5. Evaluación del pragmatismo a través de la herramienta PRECIS-2: Se calcularon los valores medios, incluyendo todos los dominios informados. Para este estudio, los revisores calificaron los dominios no informados como «en blanco» y los excluyeron de la puntuación.

6. Limitaciones auto-reportadas por los autores: Los datos de extracción se basaron en las 12 categorías de limitaciones propuestas por Álvarez et al. [80].

El formulario de extracción de datos se puso a prueba con cinco ECAs para evaluar las discrepancias de los revisores. Después de la fase piloto, el formulario de extracción de datos incorporó las sugerencias de todos los revisores para mejorar la coherencia. Finalmente, se proporcionó al equipo una guía para los revisores (**ANEXO 2**).

3.1.5 Análisis de datos

Se realizó un análisis descriptivo de las variables categóricas, donde los resultados se presentaron como frecuencias relativas y absolutas. Asimismo, se utilizaron medias y desviaciones estándar para describir variables cuantitativas. Para las variables ordinales se aplicaron medianas y rangos intercuartílicos. Se utilizó la rueda PRECIS-2 para presentar gráficamente la actitud pragmática de cada artículo. Además, las variables se compararon mediante ANOVA para variables cuantitativas, Chi-cuadrado para variables categóricas y Krustall Wallis para variables ordinales. Siguiendo investigaciones anteriores [103,104,106], empleamos una escala de 9 a 45 puntos para evaluar la puntuación total de PRECIS-2, aparte del promedio. Las calificaciones de 9 a 22 se categorizaron como 'ligeramente pragmático', de 23 a 34 como 'moderadamente pragmático' y de 35 a 44 como 'muy pragmático'. Realizamos un análisis post hoc utilizando la prueba de Scheffé para ANOVA y la prueba de Mann-Whitney para Krustall Wallis. Se realizó la prueba de Spearman para analizar la correlación entre PRECIS-2 y algunas características descriptivas. El nivel de significancia se fijó en 5% ($\alpha=0,05$). Todos los datos se analizaron utilizando el software IBM-SPSS (V26.0).

3.2 Métodos del segundo trabajo

El objetivo principal del segundo trabajo consistió en evaluar la calidad del reporte y la validez interna de los estudios de efectividad auto-denominados pragmáticos en el título y/o resumen en el campo de la terapia manual. Como objetivo secundario se examinó la relación entre el riesgo de sesgo, el reporte y el grado de pragmatismo de los estudios.

3.2.1 Criterios de elegibilidad de los estudios incluidos

Los criterios de elegibilidad de los ECAs para este trabajo son los mismos descritos en el apartado 3.1.1 ya que se utilizó la misma metodología.

3.2.2 Estrategia de búsqueda

La estrategia de búsqueda utilizada para este trabajo fue la misma descrita en el apartado 3.1.2 ya que se utilizó la misma metodología.

3.2.3 Selección de los estudios

La selección de los estudios para este trabajo fue la misma descrita en el apartado 3.1.3 ya que se utilizó la misma metodología.

3.2.4 Proceso de recopilación y extracción de datos

Dos revisores independientes extrajeron los datos de los estudios incluidos y resolvieron los desacuerdos por consenso.

La extracción de datos incluyó los siguientes elementos:

1. Calidad del reporte a través de analizar el cumplimiento de los ítems de la Guía CONSORT y la extensión para estudios pragmáticos e intervenciones no farmacológicas [97,127,128].
2. Riesgo de sesgo a través de la herramienta Cochrane RoB [129].

3.2.5 Análisis de datos

En análisis de los datos para este trabajo fue el mismo descrito en el apartado 3.1.5 ya que se utilizó la misma metodología.

3.3 Métodos del tercer trabajo

El objetivo del tercer trabajo consistió en evaluar las características tanto de diseño, reporte y riesgo de sesgo de los estudios de efectividad en el campo de la terapia manual y compararlos con las características de los estudios de efectividad autodenominados pragmáticos encontradas en los dos primeros trabajos. Esto permitió valorar cuán pragmáticos son los estudios pragmáticos en TM y valorar el uso del término “efectividad” en la comunidad científica específicamente en este campo.

3.3.1 Criterios de elegibilidad de los estudios incluidos

Los criterios de elegibilidad de los estudios son los mismos que los descritos en el apartado 3.1.1.

3.3.2 Estrategia de búsqueda

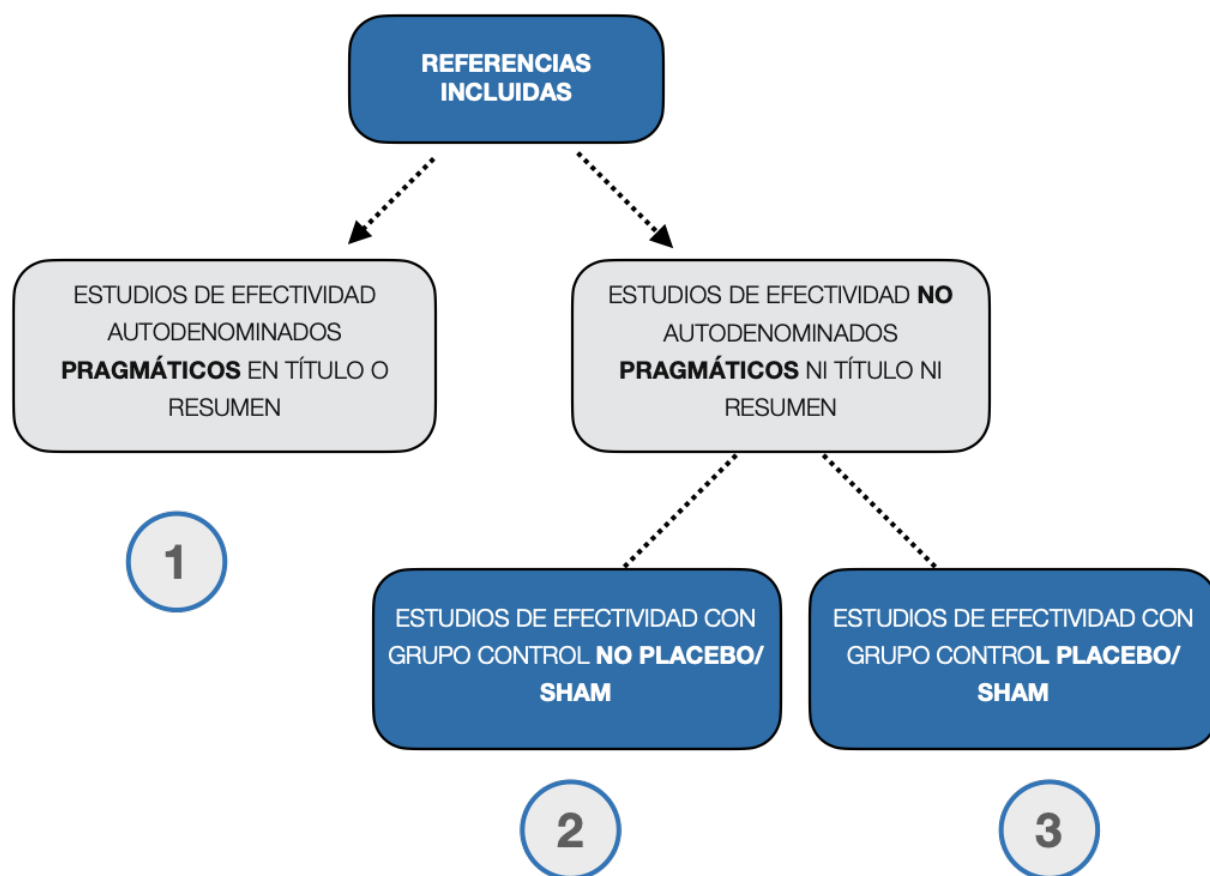
La estrategia de búsqueda utilizada para este trabajo fue la misma descrita en el apartado 3.1.2 ya que se utilizó la misma metodología.

3.3.3 Selección de los estudios

La selección de los estudios para este trabajo fue la misma descrita en el apartado 3.1.3 ya que se utilizó la misma metodología. No obstante, en este tercer trabajo, las referencias incluidas en la revisión se clasificaron inicialmente en dos grupos: en el primer grupo se incluyeron los estudios de efectividad autodenominados como "pragmáticos" en el título y/o resumen (EP), y en el segundo grupo se incluyeron los estudios de efectividad que no se autodenominaban como pragmáticos (EE). Para un análisis más detallado, dividimos el grupo de EE en estudios que utilizaban un grupo de control con placebo/sham (EEplacebo) y aquellos que emplearon un comparador activo aceptado o establecido (EEotraitervención), resultando finalmente tres grupos (**Figura 8**).

Determinamos la muestra definitiva emparejando todos los grupos con el número equivalente a los EPs recuperados, aleatorizando el número de estudios de efectividad, para equiparar los grupos, utilizando el generador aleatorio de Microsoft Excel.

Figura 8. Distribución de los estudios incluidos en la revisión



3.3.4 Proceso de recopilación y extracción de datos

El proceso de recopilación y extracción de datos fue el mismo que en los dos trabajos anteriores, descritos en el punto 3.1.4 y 3.2.4.

3.3.5 Análisis de datos

En análisis de los datos para este trabajo fue el mismo descrito en el apartado 3.1.5 ya

que se utilizó la misma metodología.

4. RESULTADOS

Los resultados de esta tesis presentada como compendio de publicaciones son los estudios detallados a continuación:

1. Roura, S., Alvarez, G., Hohenschurz-Schmidt, D., Solà, I., Núñez-Cortés, R., Bracchiglione, J., C. Fernández-Jané, Phalip, J., Gich, I., Sitjà-Rabert, M., & Urrútia, G. (2024). **Lack of pragmatic attitude of self-labelled pragmatic trials on manual therapy: a methodological review.** *BMC Medical Research Methodology*, 24(1). <https://doi.org/10.1186/s12874-024-02393-1>
2. Roura, Alvarez, Hohenschurz-Schmidt, D., Solà, Núñez-Cortés, R., Bracchiglione, Fernández-Jané, C., Phalip, Gich, Sitjà-Rabert, M., & Urrútia. (2025). **A call for improving the Internal validity and the reporting of manual therapy trials self-labelled as pragmatic: A methodological review.** *International Journal of Osteopathic Medicine: IJOM*, 100754, 100754.
3. Roura, S., Alvarez, G., Hohenschurz-Schmidt, D., Solà, I., Núñez-Cortés, R., Bracchiglione, J., Fernández-Jané, C., Phalip, J., Gich, I., Sitjà-Rabert, M., & Urrútia, G. (2025). **Alignment between the research question, design and terminology is required in manual therapy trials - a methodological study.** *Journal of Clinical Epidemiology*, 183, 111817.

A continuación, se resumen los principales resultados de los tres trabajos. Sin embargo, para una completa interpretación de los mismos, se remite al lector a las publicaciones

4.1 Resultados del primer trabajo

4.1.1 Resultados de la búsqueda

Después de eliminar los estudios duplicados, la búsqueda identificó 3553 artículos. Después de evaluar el título, el resumen y la disponibilidad del texto completo, la muestra final estuvo compuesta por 39 estudios identificados como autodenominados pragmáticos (EP) en el título y/o resumen.

4.1.2 Características principales de los estudios autodenominados pragmáticos

De los 39 EPs analizados, el 75 % (30/39) buscaba evaluar la efectividad de una intervención, el 8 % (3/39) intentaba reproducir la práctica clínica y el 15 % (6/39) combinaba ambos propósitos. El 64 % (25/39) justificó en la introducción la necesidad de un EP: el 56 % de estos (14/25) apeló a estudios de efectividad comparativa, el 44 % (11/25) a estudios mecanicistas y uno (4 %) a estudios con placebo. El 21 % (8/39) no aportó justificación y en el 15 % (6/39) no fue claro.

No se observó un aumento en las publicaciones de EPs en TM a lo largo del tiempo. Respecto al tipo de revista, el 44 % (17/39) apareció en publicaciones especializadas en TM y el 56 % (22/39) en no especializadas. El 41 % (16/39) de los estudios se reportó según CONSORT, pero solo uno declaró haber seguido la extensión para estudios pragmáticos, y apenas un estudio empleó la herramienta PRECIS-2 en el diseño.

Como variable principal, el 72 % (28/39) utilizó una PROM. La intervención más frecuente

(54 %, 21/39) fue una combinación de técnicas no protocolizadas. Entre las modalidades específicas se encontraron: técnicas manuales no especificadas (23 %, 9/39), quiropráctica (18 %, 7/39), manipulación espinal (18 %, 7/39), osteopatía (10 %, 4/39), Chuna (10 %, 4/39), reflexología (8 %, 3/39), liberación miofascial (5 %, 2/39), masaje (5 %, 2/39) y acupresión (3 %, 1/39). El comparador más habitual fue el “usual care” (33 %, 13/39), seguido de una intervención activa (31 %, 12/39); solo dos estudios (5 %) usaron placebo. Las limitaciones más citadas fueron la falta de enmascaramiento de participantes (36 %, 14/39), controles inadecuados (33 %, 13/39), dificultades para generalizar resultados (33 %, 13/39) y tamaño muestral reducido (28 %, 11/39).

4.1.3 Evaluación retrospectiva del pragmatismo con la herramienta PRECIS-2

Los estudios incluidos en nuestra muestra obtuvieron una puntuación promedio de 3,5 (SD = 0,6) en todos los dominios de PRECIS-2. La Figura 2 (de la publicación respectiva) muestra la representación de la rueda PRECIS-2 en relación a la actitud pragmática de los estudios de nuestra muestra. La puntuación total de PRECIS-2 (que oscila entre 9 y 45) mostró que el 61% (24/39) de los estudios fueron moderadamente pragmáticos, el 36% (14/39) muy pragmáticos y el 3% (1/37) ligeramente pragmáticos. La puntuación sobre el pragmatismo no aumentó con el tiempo yendo desde una media de 3,77 (SD = 0,49) antes de 2012 a 3,67 (SD = 0,53) después de 2018. En general, hubo una disminución en el puntaje promedio durante el período 2012-2017; después de eso, se recuperaron, pero no significativamente en 8 de 9 dominios. Por el contrario, hubo un aumento estadísticamente significativo en el Dominio 5 ($p=0,002$). Este valor fue confirmado con una prueba no paramétrica mostrando el mismo incremento estadístico ($p=0,006$).

El Dominio 8 (variable primaria) se puntuó pragmático (eso significa que puntuó 4 o 5 con la herramienta PRECIS-2) en prácticamente toda la muestra. Los Dominios 5 (flexibilidad de la intervención), 7 (seguimiento) y 9 (análisis primario) se consideraron pragmáticos en

56

el 60% de la muestra y el Dominio 1 (elegibilidad) ,2 (reclutamiento) , 3 (entorno)y 4 (organización de la intervención) se consideraron pragmáticos sólo en la mitad de la muestra (54%,51%, 54%, 52%). El resto se consideraron explicativos, igual de explicativos que pragmáticos (puntuación 3 en la herramienta PRECIS-2).

4.1.4 Primera publicación

Roura, S., Alvarez, G., Hohenschurz-Schmidt, D., Solà, I., Núñez-Cortés, R., Bracchiglione, J., C. Fernández-Jané, Phalip, J., Gich, I., Sitjà-Rabert, M., & Urrútia, G. (2024). **Lack of pragmatic attitude of self-labelled pragmatic trials on manual therapy: a methodological review.** *BMC Medical Research Methodology*, 24(1). <https://doi.org/10.1186/s12874-024-02393-1>

RESEARCH

Open Access



Lack of pragmatic attitude of self-labelled pragmatic trials on manual therapy: a methodological review

S. Roura¹, G. Alvarez^{2,3*}, D. Hohenschurz-Schmidt^{4,5}, I. Solà^{2,6,7}, R. Núñez-Cortés⁸, J. Bracchiglione^{2,6,9}, C. Fernández-Jané^{3,10,11}, J. Phalip^{12,13}, I. Gich^{6,14}, M. Sitjà-Rabert^{3,10} and G. Urrútia^{2,6,7}

Abstract

Background Pragmatic randomized controlled trials are getting more interest to improve trials' external validity. This study aimed to assess how pragmatic the design of the self-labelled pragmatic randomised controlled trials in the manual therapy field is.

Methods We searched MEDLINE and the Cochrane Central Register of Controlled Trials for self-labelled pragmatic randomised controlled trials in the manual therapy field until January 2024 were included. Two independent reviewers collected and extracted data related to the intention of the trial, the rationale for the intervention, and specific features of the trial and performed an assessment using the PRECIS-2 tool.

Results Of 39 self-labelled pragmatic trials, the mean PRECIS-2 score was 3.5 (SD: 0.6). Choice of outcome measures, how the interventions were performed, the follow-up of the participants and how all the available data were included in the statistical analysis were the domains rated as most 'pragmatic'. Participants' eligibility, recruitment, and setting obtained lower scores. Less than 25% of the trials claimed that the aim was to investigate an intervention under real-world conditions and to make clinical decisions about its effectiveness. In the 21% of the sample the authors described neither the proof-of-concept of the intervention nor the state of previous studies addressing related research questions.

Conclusions Self-labelled pragmatic randomised controlled trials showed a moderately pragmatic attitude. Beyond the label 'pragmatic', the description of the intention of the trial and the context of every PRECIS-2 domain is crucial to understanding the real pragmatism of a trial.

Keywords Pragmatic clinical trials, Musculoskeletal manipulations, Randomized controlled trials, Manual therapy, Generalizability, PRECIS-2

*Correspondence:

G. Alvarez
gerardalv@gmail.com

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Background

Medical research is often criticised for its low resemblance to real clinical settings [1, 2]. Moreover, there is a large gap between the amount of medical research available and that useful for healthcare practitioners [3]. In that sense, there is an increasing interest in trials designed with a pragmatic attitude to respond to the currently poor generalisability of many studies [1, 4, 5].

Pragmatic trials were introduced more than 50 years ago by Schwartz and Lellouch [6], who proposed classifying trials according to their “attitude” as (a) trials whose final goal is to examine the causal relationship between an intervention and a physiological effect (explanatory/efficacy) or (b) trials that aim to inform clinicians or health policy decision-making directly (pragmatic/effectiveness). Explanatory trials assess how a specific intervention impacts participants optimally, often using placebos or active comparators. They focus on measurable symptoms or markers and aim to minimise participant variability. In contrast, pragmatic trials evaluate interventions in real-world clinical settings to ensure broader applicability and effectiveness across diverse patient populations and treatment settings, requiring larger, more flexible designs [7]. It is essential to highlight that the pragmatic term refers to an attitude rather than a characteristic of the study and that a continuum exists from explanatory to pragmatic [7]. Therefore, some efficacy trials can have pragmatic features to increase their external validity, and pragmatic randomised controlled trials (pRCT) can use explanatory elements to ensure the trial’s internal validity [6, 8]. The terms commonly used to show that a given study has a pragmatic attitude are “pragmatic”, “naturalistic”, and “effectiveness” (rather than “efficacy”) [6, 9–11]. Pragmatic trials have been used in pharmacological research in phase IV trials [12]. However, the development of pharmacological and non-pharmacological interventions differs substantially because non-drug interventions are not as commonly regulated as drugs are (e.g. through the Food and Drug Administration or the European Medicines Agency). Arguably, this situation encourages trialists to attempt a pragmatic approach directly but using explanatory designs, concluding with recommendations of treatment (pragmatic attitude) regardless of the absence of high-quality efficacy trials (this was also observed by Zwarestein et al.) [13]. Although pragmatism is possible in both early and late-stage development trials, caution is required because, without efficacy evidence, results can be strongly biased toward a positive outcome [1, 14–16]. This is especially concerning for therapies not formally recognised and that lack high-quality efficacy trials [1, 16].

Various health professionals use manual therapy (MT) interventions to treat pain and disability [17–19]. Manual therapies are considered complex interventions (CI), and

like many non-pharmacological interventions, it might be challenging to evaluate them scientifically [20–22]. Some frameworks have been proposed to assess CIs more adequately [23–25]. Within the proposals for improving the applicability of RCTs [1, 21, 26], pragmatic randomised controlled trials (pRCT) have been proposed as a suitable approach in the field of MTs [17, 21]; for example, using more heterogeneous populations with similar comorbidities to those patients seen habitually in practice or using non-protocolised interventions to respect the person-centred approach required in MT treatments [27]. However, a recent systematic review revealed that MT research lacks applicability to real practice [28]. Despite the increasing interest in designing pragmatic trials, some doubts arise about the real pragmatic attitude of those trials self-labelled “pragmatic” [9, 10, 29].

To date, it is unknown how pragmatic trials are used among MT researchers. Therefore, this review aims to assess, through the published article information, how pragmatic the design of the self-labelled pRCTs in the MT field is. Secondly, we aim to examine whether and how authors report the intention of the trial and how design choices are justified.

Methods

Protocol registration

We conducted a methodological review of pragmatic randomised controlled trials on manual therapy interventions and reported its findings according to PRISMA guidelines [30] (Supplementary file 1). We prospectively registered the protocol on the Open Science Framework (DOI <https://doi.org/10.17605/OSF.IO/WKEPZ>).

Eligibility criteria

To be eligible for inclusion in the review, RCTs had to meet specific criteria regarding the methods used and the experimental intervention. Concerning the methods, we included reports of RCTs that used the terms “pragmatic” or “naturalistic” (about the methodological design) either in the title or the abstract [10, 11]. Regarding the intervention used, we followed the criteria of previous research in this field [17], requiring eligible references to include either one manual technique or a combination of them. These techniques included soft tissue techniques, joint mobilisations or manipulations, massage, myofascial release, nerve manipulation, strain/counterstrain, and acupuncture. We applied no restrictions regarding population, comparators or outcome measures. Exclusion criteria included experimental interventions delivered through tools, devices (electrotherapy, kinesiotaping, dry needling, acupuncture), drugs, active exercises or a combination of therapies without an MT intervention. We also excluded articles not written in English, protocols and poster/conference presentations.

Search strategy

We searched MEDLINE and the Cochrane Central Register of Controlled Trials from their inception to January 2024. The search strategy combined controlled vocabulary related to the field of MT and the design of interest (Supplementary file 2).

Study selection

The records retrieved from the search were imported to Rayyan software (www.rayyan.ai) [31]. After deduplication, two reviewers (SR, GA) independently screened the references based on title and abstract, resolving disagreements through discussion.

Data collection process

Two independent reviewers (SR and an additional reviewer among GA, RN, JB, DH, CF, and JP) collected data from the included studies, resolving disagreements by consensus. Data was collected from an extraction data form designed by 3 authors (SR, GA, GU). Before the pilot of the data extraction form, the reviewers discussed and received training from Dr Kirsty Loudon (developer of the PRECIS-2 tool [32]); note that PRECIS-2 tool will be explained below. After piloting five trials, the data extraction form incorporated all reviewers' suggestions to improve consistency. Finally, a guideline for reviewers was provided to the team (Supplementary file 3).

Data items

Recent literature has argued that a part of quantitatively assessing the degree of pragmatism of RCTs is important to understand the intent of the trial and how the design relates to this intent [4]. Authors typically use it to refer to the trial as a whole (instead of their intention) and often do not provide adequate supporting justification for its use [29]. For this reason, our data extraction form included the following:

1. Bibliometric identification elements and specific trial characteristics.
2. The intention of the trial given by the authors: assessed using the following three qualitative categories pre-established by the reviewer team, aimed at how authors reported if their aim was (I) examining the effectiveness of an intervention (only reporting how participants respond to treatment, examining and estimating effects/changes); (II) resemble clinical practice in terms of populations included, setting, or intervention delivery; and (III) resemble clinical practice and informing clinical decision-making [4, 29].
3. The rationale of the intervention given by the authors: This information was screened in the introduction section of the trials. It refers to

the existence of any report of proof-of-concept information related to the intervention. Reviewers pre-established three types of rationale, reporting: (I) the existence of mechanistic experiments explaining the mechanisms of action of the intervention; (II) the existence of sham-controlled studies testing the intervention against a placebo; and (III) comparative effectiveness data about the intervention [33].

4. Experimental and control interventions: The research team used a categorisation process to distinguish between the experimental and control interventions. This was done to assess the extent to which the intervention differed from usual care. As a patient-centred intervention, a pragmatic approach to MT requires individual tailoring for each person and also adaptation of different techniques as the condition progresses [27]. Part of the research team were experts in MT and established a classification of the intervention according to its ability to resemble clinical practice. Box 1 refers to pre-established categories to classify the experimental and control interventions. Experimental interventions were divided into those that were protocolised due to the requirements of the study and those that led to the therapist's decisions depending on the patient's needs.
5. PRECIS-2 tool assessment: It consists of nine domains based on trial design decisions (Eligibility criteria, Recruitment, Setting, Organisation, Flexibility delivery, Flexibility adherence, Follow-up, Primary outcome and Primary analysis) [32]. Every domain is scored on a 5-point Likert-type scale, with a score of 1 being considered very explanatory and a score of 5 being considered very pragmatic. Scores for each domain are represented on a radar chart (PRECIS-2 wheel). Trials with more explanatory scores show a narrower wheel towards the centre, and trials containing more pragmatic scores have a wider wheel towards the periphery. Mean values were calculated, including all reported domains [32]. For this study, reviewers rated the non-reported domains "blank" and excluded them from the score.
6. Limitations reported by the authors: The extraction data was based on the 12 limitations categories proposed by Alvarez et al. [34]

Data analysis

We performed a descriptive analysis of the categorical variables, presenting the results as relative and absolute frequencies. Quantitative variables were described with means and standard deviations. Median and interquartile ranges defined ordinal variables. Each article's pragmatic attitude was graphically presented using the PRECIS-2

BOX 1 Pre-established experimental and control intervention categories

Experimental intervention categories	
Combination of non-protocolised techniques	The use of a combination of MT techniques already in use in clinical practice but chosen by the practitioner depending on the participant's needs.
Isolated non-protocolised technique	The practitioner chooses a single MT technique depending on the participant's needs.
Protocol of a combination of techniques	Use of a combination of MT techniques standardised by the trial protocol
Combination of non-protocolised therapies	Use of a combination of different modalities of treatment based on the needs of patients, i.e. MT combined with exercises
Isolated protocolised technique	Use of a single MT technique standardised by the trial protocol
Protocol of a combination of therapies	Use of a combination of therapies standardised to all patients by the trial protocol
Control intervention categories	
Active intervention	Use of another accepted or established intervention that is not considered beyond the label 'usual care.'
Placebo/sham	Use of a supposedly inactive intervention that looks like the drug or treatment being tested
Usual care	The intervention expected to be received as part of normal clinical care
No intervention	A group not receiving any intervention

wheel. Variables were compared using ANOVA for quantitative variables, Chi-square for categorical variables and Kruskal Wallis for ordinal variables. According to previous literature [35–37], we used a scale ranging from 9 to 45 to assess the PRECIS-2 total score, a part of the PRECIS average. A score of 9–22 was considered slightly pragmatic, 23–34 moderately pragmatic, and ≥ 35 very pragmatic. The significance level was 5% ($\alpha=0.05$). All data were analysed using the software IBM-SPSS (V26.0).

Results

Study selection

After removing duplicated references, the search yielded 3553 unique references. Following the screening process, we included 39 self-labelled MT pRCTs [38–76]. Figure 1 provides the PRISMA diagram.

Study characteristics

Table 1 summarises the main characteristics of our sample. Three-quarters (27/39) of the pRCTs aimed to examine the effectiveness of an intervention, 8% (3/39) aimed at resembling clinical practice, and 23% (9/39) pursued both objectives. Two-thirds (25/39) of the studies provided the rationale for assessing the intervention with a pRCT in their introduction (of which 15 trials (60%) reported comparative effectiveness data, 11 trials (44%) reported mechanistic experiments, and one

(4%) reported sham-controlled studies). In comparison, 21% (8/39) of the sample provided no rationale, and 15% (6/39) were unclear. Details and direct quotes from the authors about the intention, the rationale of the trial, and how authors justify them can be found in supplementary file 4.

Although this analysis was not pre-planned, our results showed no increase in MT pRCT publications over time (Fig. 2). Regarding the type of journal, 44% (17/39) of the studies were published in specialised journals in the MT field, and 56% (22/39) in journals non-specialised in MT. One-third (16/39) of the trials were reported according to the CONSORT statement, but only one study [44] reported their findings adhering to the CONSORT extension for pragmatic trials. Only one study reported using the PRECIS-2 tool in its design phase [50].

The primary outcome was a patient-reported outcome measure (PROM) in 72% of the sample (28/39). The intervention most frequently used was a combination of non-protocolised techniques (54%, 21/39). The modality of MT used was an unspecified combination of manual techniques (23%, 9/39), chiropractic treatment (18%, 7/39), spinal manipulation (18%, 7/39), osteopathy (10%, 4/39), Chuna treatment (10%, 4/39), reflexology (8%, 3/39), myofascial release (5%, 2/39), massage therapy (5%, 2/39), and acupressure (3%, 1/39). The most common control intervention was usual care (33%, 13/39) followed by an active intervention (31%, 12/39), whereas a placebo was used in only two studies (5%, 2/39). The most common self-reported limitation by the authors was lack of blinding of the participants (36%, 14/39), inadequate control (no-treatment control or a sham control group to assess the natural progression of the pathology) (33%, 13/39) and compromised generalisability (33%, 13/39), followed by sample size (28%, 11/39).

PRECIS assessment

The studies included in our sample scored an average of 3.5 (SD=0.6) across all PRECIS-2 domains (median 3.5; IQR: 3.1–3.9). Figure 3 shows the PRECIS-2 wheel representation of the pragmatic attitude from trials in our sample. We show PRECIS-2 wheels for each individual included trial in Supplementary file 5. Table 2 presents data for individual domains. The PRECIS-2 total score (ranging from 9 to 45) showed that 61% (24/39) of the studies were moderately pragmatic, 36% (14/39) highly pragmatic, and 3% (1/39) slightly pragmatic. The pragmatism score did not increase over time from a mean of 3.68 (SD=0.49) before 2012 to 3.67 (SD=0.57) after 2018 (Fig. 2). In general, there was a decrease in the mean score during the period 2012–2017; after that, they recovered but not significantly in 8 of 9 domains (D1 $p=0.250$; D2 $p=0.157$; D3 $p=0.729$; D4 $p=0.958$; D6 $p=0.654$; D7 $p=0.838$; D8 $p=0.271$ y D9 $p=0.480$). On

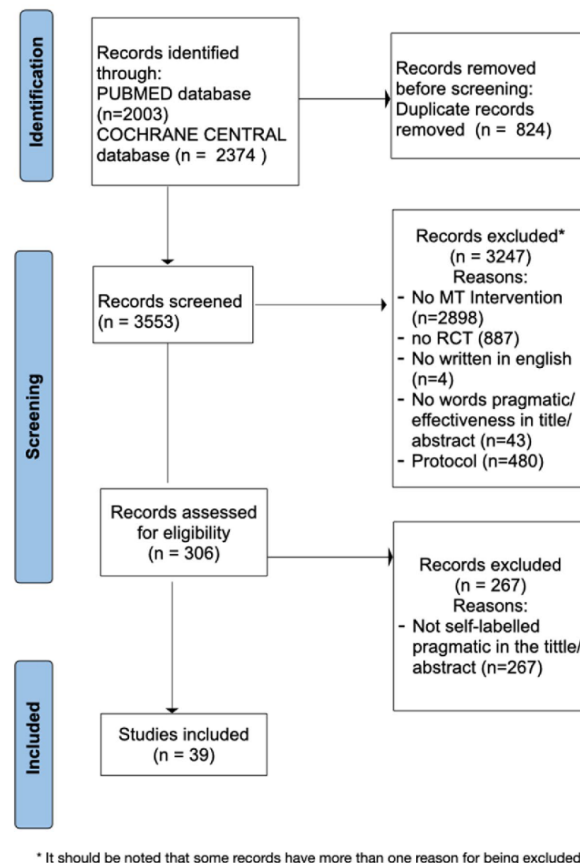


Fig. 1 PRISMA flow diagram [30]

the contrary, there was a statistically significant increase in Domain 5 ($p=0.002$). This value was confirmed with a non-parametric test showing the same statistical increase ($p=0.006$).

The number of studies rated for each domain was: Eligibility 39/39, Recruitment 37/39, Setting 38/39, Organisation 39/39, Flexibility delivery 39/39, Flexibility adherence 32/39, Follow-up 39/39, Primary Outcome 39/39, Primary analysis 38/39.

Figure 4 shows the PRECIS-2 scores for each domain. Domain 8 was rated as pragmatic (it means that they scored 4 or 5 in the PRECIS-2 score) by almost all the sample. Domains 5, 7 and 9 were considered pragmatic by 60% of the sample and Domains 1, 2, 3, and 4 were rated pragmatic only in half of the sample (54%, 51%,

54%, 52%), (Fig. 5). The rest were explanatory or equally explanatory and pragmatic.

Discussion

We aimed to describe the characteristics related to the design of a cohort of self-labelled pRCTs in MT according to their pragmatic attitude. We obtained three main findings. First, MT pRCTs are not becoming more common despite the increased attention on this type of design. Second, our findings revealed that pRCTs in MT have a moderate pragmatic attitude. However, the label 'pragmatic' could be questioned in those trials using highly explanatory design features like blinding, unicentric setting, or a placebo as a control intervention. Lastly, the authors of our sample of pRCTs did not adequately

Table 1 Main characteristics of pRCTs in MT

Number of participants (Mean)		169,23 (SD = 150.7)
N of participants		Per cent (n**)
< 50		15% (6)
51–100		23% (9)
101–200		38% (15)
> 201		23% (9)
Experimental intervention		
Combination of non-protocolised techniques		54% (21)
Isolated non-protocolised technique		15% (6)
Protocol of a combination of techniques		13% (5)
Combination of non-protocolised therapies		10% (4)
Protocol of an isolated technique		5% (2)
Protocol of a combination of therapies		2% (1)
Control intervention		
2 arms	test treatment vs. other active intervention	31% (12)
	test treatment vs. placebo	2% (1)
	test treatment vs. usual care	33% (13)
	test treatment vs. no intervention	13% (5)
3 arms	test treatment vs. 2 other active interventions	5% (2)
	test treatment vs. 1 other intervention and 1 placebo	2% (1)
	test treatment vs. 1 intervention and 1 usual care	10% (4)
	test treatment vs. 1 other intervention and 1 no intervention	2% (1)
Blinding (Yes)		
Participants		10% (4)
Therapists		0% (0)
Outcome assessors		20% (8)
Statistician		26% (10)
Setting		
Multicentric		56% (22)
Unicentric		33% (13)
Unclear report		10% (4)
Follow-up		
No follow-up		18% (7)
< 2 weeks		0% (0)
2–4 weeks		2% (1)
4–12 weeks		10% (4)
3–6 months		18% (7)
6–12 months		20% (8)
> 1 year		26% (10)
Individualised*		5% (2)

*Two studies used an individualised follow-up (trials assessing how many weeks patients remained pain-free)

**N of studies = 39

justify why the trial aimed to be pragmatic and which domains most contributed to this pragmatism.

Although research in the MT field has increased in the last decades [77–79], pRCTs are not so common. Only 39 papers assessing an MT intervention were identified as self-labelled pragmatic from inception to 2024. Our results indicate that the number of MT pRCTs has not increased in recent years, and the PRECIS-2 mean score has also not increased over time. Nevertheless, despite our comprehensive analysis of the available literature, the limited number of articles prevents us from drawing any definitive conclusions. These results are contrary to studies assessing pragmatism in critical care [80], pain [5] and cardiovascular interventions [81], wherein the mean PRECIS score increased over time. Only Domain 5 (flexibility intervention) increased after 2018, primarily due to the significant decrease in the mean PRECIS scores between 2012 and 2018. It might be interesting to explore the knowledge about non-pharmacological pRCTs among researchers to understand this lack of increase over time. It might be interesting to assess the necessity of including specific training on this topic in the academic stages.

The mean score of our sample in the retrospective PRECIS-2 tool assessment was 3.5 (SD = 0.6), indicating a moderate level of pragmatism of MT pRCTs. Those are similar ratings to a recent systematic review of nursing interventions and those found in pain interventions and other fields [5, 9, 10, 29, 35, 80, 81]. The moderate level of pragmatism found in our review was highly influenced by the domains “flexibility intervention”, “follow-up”, “outcome measure”, and “primary analysis”. Those domains were rated as pragmatic (4 or 5 on the Likert scale) in more than 60% of our assessed trials. Using an intention to treat analysis is a trait highly pragmatic, and most of the reviews assessing the pragmatism of trials in other fields found this item as the most pragmatic in their PRECIS-2 tool assessment [5, 35, 80, 81].

Considering the extensive debate about the appropriateness of self-labelled pragmatic trials using explanatory features such as blinding, unicentric setting and use of placebo controls [12, 13, 82–85, 90], we might exclude almost half of our sample (19/37). However, as there is no consensus on that matter, we decided to assess the whole sample and to further contribute to the debate as other authors did in other fields [5, 9, 29, 35].

Using a placebo as a comparator is considered a very explanatory feature. Some authors argue that it is hardly part of real-world care, and its use suggests an explanatory nature of the trial [12, 82–84]. PRECIS-2 authors defend that there are situations in which a trial could include a placebo control and still provide useful information to decision-making (e.g. to mitigate the subjectiveness of patient-reported outcomes in a comparative

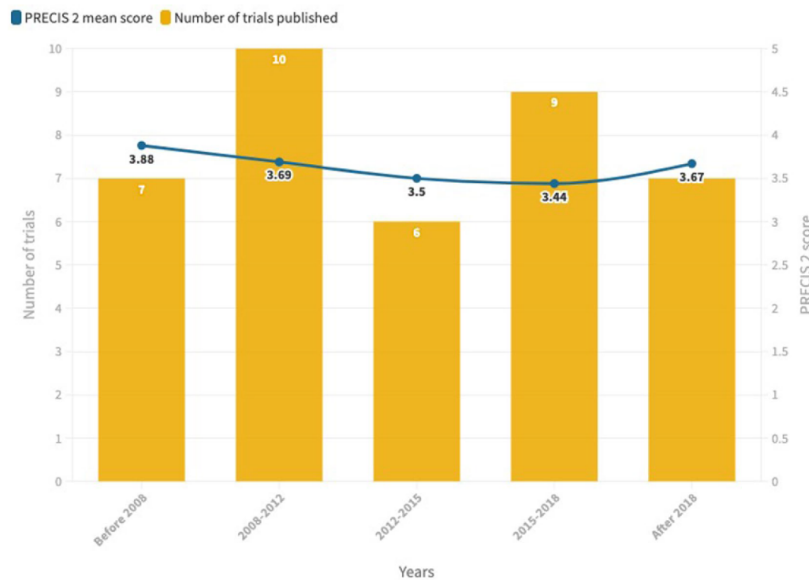


Fig. 2 The number of published pRCTs and mean PRECIS 2 score over time

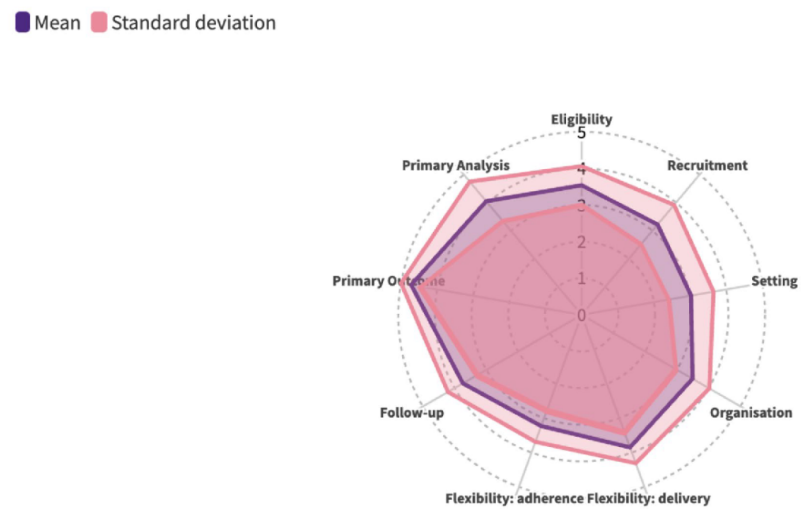


Fig. 3 Wheel diagram showing the mean PRECIS-2 score of the included studies

effectiveness trial of two accepted treatments) [7, 85]. We wonder if including an appropriate placebo/sham comparator could be interesting for understanding the real effect of the intervention, which is difficult to achieve in an efficacy study under non-real-world conditions (where all the contextual factors part of the patient-therapist encounter, are not present [86–89]). In our sample, only one study used a placebo control [39], and another

Table 2 Mean scores of each PRECIS-2 domain over time. For statistical reasons, years were divided into three periods (before 2012, 2012–2018, and after 2018)

	Mean (SD)			
	Overall N = 39/39	Before 2012 N = 17	2012–2018 N = 15	After 2018 N = 7
Domain 1: Eligibility	3.54 (1.04)	3.82 (0.72)	3.33 (1.11)	3.29 (1.5)
Domain 2: Recruitment	3.22 (1.40)	3.65 (1.37)	2.79 (1.31)	3 (1.6)
Domain 3: Setting	3.03 (1.24)	3.06 (1.14)	3.07 (1.38)	2.83 (1.33)
Domain 4: Organization	3.49 (1.07)	3.53 (1.01)	3.47 (1.30)	3.43 (0.80)
Domain 5: Flexibility delivery	3.85 (0.93)	4.12 (0.70)	3.47 (1.06)	4.00 (1.00)
Domain 6: Flexibility adherence	3.23 (0.92)	3.15 (1.14)	3.23 (0.83)	3.40 (0.55)
Domain 7: Follow-up	3.74 (0.93)	3.82 (0.95)	3.73 (1.03)	3.57 (0.78)
Domain 8: Primary Outcome	4.72 (0.64)	3.82 (1.51)	4.36 (1.15)	4.00 (1.73)
Domain 9: Primary analysis	4.05 (1.41)	3.82 (1.51)	4.30 (1.25)	4.44 (1.13)

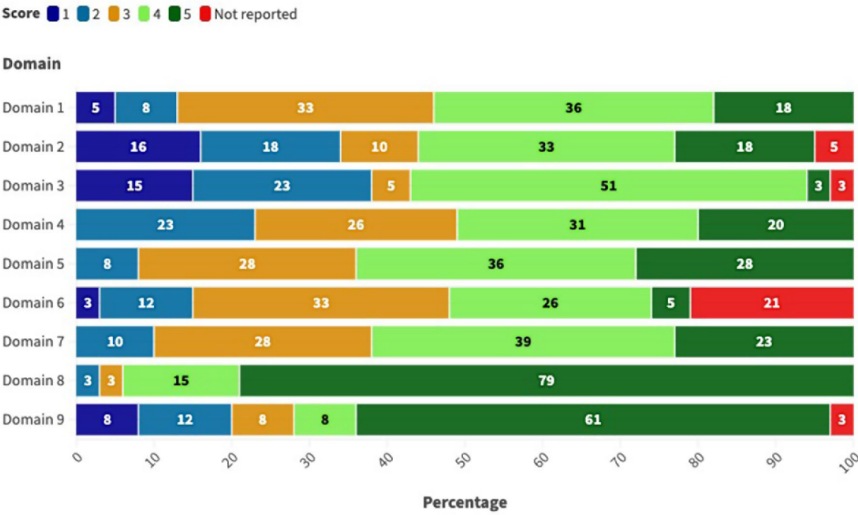


Fig. 4 PRECIS-2 scores of the included studies

included a placebo control in a 3-armed trial [57]. In the latter, Mafetoni et al. showed that although the effect of the intervention was statistically larger than that in the control and placebo groups, the placebo group showed a larger effect than the control group [57]. Non-pharmacological interventions do not follow the same developing processes as pharmacological interventions do. This fact entails that effectiveness trials are often performed to assess interventions without performing previous efficacy trials. Further research should be conducted to contribute to this rich debate about using a placebo in pRCTs in the MT field [12, 84, 90].

Another controversial point is the use of single centres, as there is a debate about whether they should be considered pragmatic because of the arguably limited

generalisability of their results [82, 84, 85, 90]. In our sample, 33% of the reports were performed in a single-centre setting.

Only one trial reported using the PRECIS-2 tool to guide where the trial falls in the explanatory-pragmatic continuum [50]. It is worth mentioning that this trial scored higher on average (4.1) than the rest. Other authors also noted this limited reporting of using a design tool [35, 80, 81].

Recently, it has been pointed out that, when assessing pragmatic attitude in a trial, special attention must be paid to detecting its intention besides awarding a quantitative score [4, 82]. By knowing the trial's intention, some domains can be expected to be more pragmatic than others. Some authors stated that the pragmatic attitude

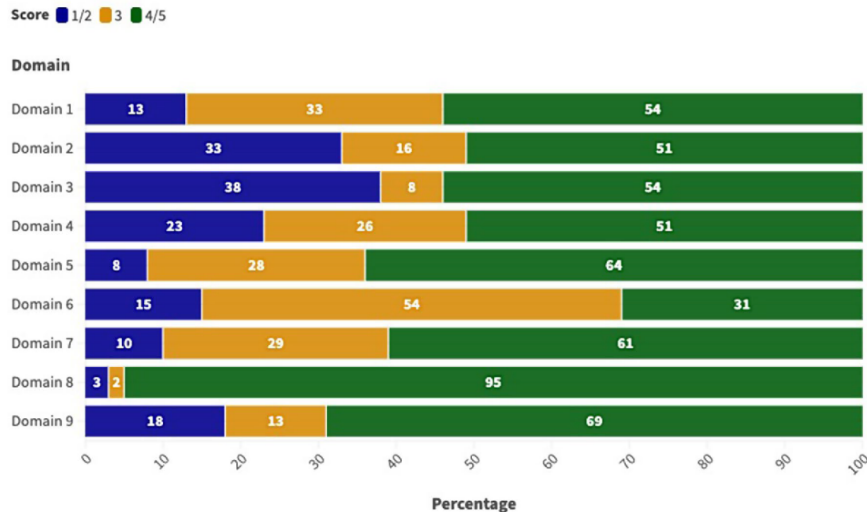


Fig. 5 Percentage of studies categorised as pragmatic (4/5) or explanatory (1/2)

is domain-specific and that PRECIS-2 ratings require context [5]. In fact, one of the fundamental principles of external validity is that context matters [3, 91]; it can help to understand the trial's attitude better, avoiding dichotomisation into pragmatic and explanatory [4]. In our sample, less than 25% of the trials reported that the aim was to investigate an intervention under real-world conditions and to make clinical decisions about its effectiveness, thus not identifying an intention aligned with a pragmatic attitude. So, almost 75% of our assessed trials did not adequately justify or discuss the claimed pragmatism. Likewise, Janiaud et al. assessed the pragmatic attitude of self-labelled trials in different fields, finding the same in 45% of the studies [9].

Both efficacy and effectiveness trials are necessary but address different research questions [28]. A potential question is whether there are already enough efficacy trials to explore further an intervention's effects in a more world-real setting. The authors should be expected to screen previous literature to appraise what is already known about a given intervention before attempting a pRCTs. Ultimately, a sound background should be provided alongside the rationale for attempting a pragmatic trial [92]. Such rationale might include existing mechanistic experiments and placebo- or sham-controlled efficacy trials, suggesting that the intervention has the desired physiological effects and efficacy [93]. Instead, in the 21% of the trials we reviewed, the authors described neither the proof-of-concept of the intervention nor the state of previous studies addressing related research

questions (rationale of the intervention). Furthermore, when the rationale was given, it was mainly constructed from comparative effectiveness trials. These results align with a similar review assessing self-labelled trials' pragmatic attitudes [29]. It found that 71.3% of the sample did not justify why a pragmatic trial was selected instead of an explanatory one [29]. This should be considered fundamental when designing a pRCT [1].

Although some authors declared that pRCTs might be an option for non-regulated interventions in early and late development trials [1, 10], we argue that researchers need to appraise the research field as a whole and determine whether there is sufficient evidence of the efficacy of a given intervention before assessing it in a real clinical environment. When efficacy trials have not sufficiently demonstrated the effects of an intervention, spending resources on a more explanatory design might be preferable to pRCT [93].

Limitations

Our review has several potential limitations. Although the reviewer team included experts in the MT field and expert methodologists, providing a multidisciplinary approach to data extraction, this might introduce some heterogeneity to the data extraction process that could have affected the results. Also, a deep knowledge of each intervention assessed is required to properly assess the pragmatism of a given trial. Other authors have reported similar difficulties [4, 5, 10, 94]. To mitigate the impact of these limitations, we piloted the extraction data form

and extensively discussed the ratings beforehand. Also, we paired all reviewers with the IP to discuss the scores and minimise subjectivity. Although a third party was not required to solve discrepancies, frequent discussions between reviewers were needed to conclude a reliable score. The small sample in this review might not be considered relevant and may not represent all the MT modalities. However, it included all pRCTs published in the MT field from inception to January 2024. Finally, it is possible that pragmatic trials may not always label themselves as such, particularly in the title and abstract. This may result in the omission of potentially eligible trials. However, this review aimed to assess those studies self-labelled pragmatic in the title and/or abstract.

Conclusions

This study highlights how pragmatism is represented in self-labelled MT pRCTs. Overall, pRCTs showed a moderately pragmatic attitude that has not increased over time. Domains showing higher rates of pragmatism were “flexibility intervention”, “follow-up”, “outcome measure”, and “primary analysis”. Beyond the label ‘pragmatic’, the description of the intention of the trial and the context of every PRECIS-2 domain is crucial to understanding the genuine pragmatism of a trial.

Abbreviations

PRECIS	Pragmatic-explanatory continuum indicatory summary
pRCT	Pragmatic randomised controlled trials
MT	Manual therapy
PROM	Patient-reported outcome measure

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12874-024-02393-1>.

Supplementary Material 1
Supplementary Material 2
Supplementary Material 3
Supplementary Material 4
Supplementary Material 5

Acknowledgements

Sònia Roura is a PhD student in Biomedical Research Methodology and Public Health in the Medical Department of the Universitat Autònoma de Barcelona, Barcelona, Spain. We appreciate the support and dedication of Dr Kirsty Loudon in the early stages of the project, teaching and guiding us with the PRECIS-2 tool. We would like to thank Roberto Acosta for his support in the first stages of the project.

Author contributions

SR: Conceptualization, Methodology, Investigation, Writing - original draft, Writing - review & editing. GA: Conceptualization, Methodology, Investigation, Writing - review & editing. DHS: Investigation, Writing - review & editing. IS: Supervision, Investigation, Writing - review & editing. RNC: Methodology, Investigation, Writing - review & editing. JB: Methodology, Investigation, Writing - review & editing. CFJ: Investigation. Jules Phalip: Investigation. IG: Formal analysis, Writing - review & editing. MSR: Supervision, Writing - review

& editing. GU: Conceptualization, Methodology, Supervision, Writing - review & editing.

Funding

SR was granted from Registro de Osteópatas de España. This organisation was not involved in the study design, data collection, analyses and interpretation of the data, and report writing.

Data availability

All data generated or analysed during this study are included in this published article [and its supplementary information files].

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Biomedical Research Methodology and Public Health in the Medical Department of the Universitat Autònoma de Barcelona, Barcelona, Spain

²Iberoamerican Cochrane Centre – Biomedical Research Institute Sant Pau, IIB Sant Pau, Barcelona, Spain

³Department of Physical Therapy, Faculty of Health Science Blanquerna, Ramon Llull University, Barcelona, Spain

⁴Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, London, United Kingdom

⁵UCO School of Osteopathy, Health Sciences University, London, United Kingdom

⁶Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP), Madrid, Spain

⁷Universitat Autònoma de Barcelona, Bellaterra, Spain

⁸Department of Physical Therapy, Faculty of Medicine, University of Chile, Santiago, Chile

⁹Interdisciplinary Centre for Health Studies (CIESAL), Universidad de Valparaíso, Viña del Mar, Chile

¹⁰Global Research on Wellbeing (GRoW) Research Group, Ramon Llull University, Barcelona, Spain

¹¹Tecnocampus, Universitat Pompeu Fabra, Mataró-Maresme, Barcelona, Spain

¹²Institut ANALGESIA, Faculté de Médecine, Université Clermont Auvergne, Clermont-Ferrand, France

¹³CHU Clermont-Ferrand, Inserm 1107 Neuro-Dol, Service de pharmacologie médicale, Université Clermont Auvergne, Clermont-Ferrand, France

¹⁴Institut d'Investigació Biomèdica Sant Pau (IIB SANT PAU), Barcelona, Spain

Received: 23 October 2023 / Accepted: 29 October 2024

Published online: 11 November 2024

References

1. Ford I, Norrie J. Pragmatic trials. *N Engl J Med*. 2016;375:454–63.
2. Greenhalgh T, Howick J, Maskrey N. Evidence Based Medicine Renaissance Group. Evidence based medicine: a movement in crisis? *BMJ*. 2014;348:g3725.
3. Huebschmann AG, Leavitt IM, Glasgow RE. Making Health Research Matter: a call to increase attention to External Validity. *Annu Rev Public Health*. 2019;40:45–63.
4. Nicholls SG, Zwarenstein M, Hey SP, Giraudeau B, Campbell MK, Taljaard M. The importance of decision intent within descriptions of pragmatic trials. *J Clin Epidemiol*. 2020;125:30–7.
5. Hohenschurz-Schmidt D, Kleykamp BA, Draper-Rodi J, Vollert J, Chan J, Ferguson M et al. Pragmatic trials of pain therapies: a systematic review of methods. *Pain*. 2021.

6. Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. *J Chronic Dis*. 1967;20:637–48.
7. Thorpe KE, Zwarenstein M, Oxman AD, Treweek S, Furlberg CD, Altman DG, et al. A pragmatic–explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *J Clin Epidemiol*. 2009;62:464–75.
8. Treweek S, Zwarenstein M. Making trials matter: pragmatic and explanatory trials and the problem of applicability. *Trials*. 2009;10.
9. Janiaud P, Dal-Ré R, Ioannidis JPA. Assessment of Pragmatism in recently published randomized clinical trials. *JAMA Intern Med*. 2018;178:1278.
10. Dal-Ré R, Janiaud P, Ioannidis JPA. Real-world evidence: how pragmatic are randomized controlled trials labeled as pragmatic? *BMC Med*. 2018;16:49.
11. Sacristán JA, Dilla T. Pragmatic trials revisited: applicability is about individualization. *J Clin Epidemiol*. 2018;99:164–6.
12. Dal-Ré R. The PRECIS-2 tool seems not to be useful to discriminate the degree of pragmatism of medicine masked trials from that of open-label trials. *Eur J Clin Pharmacol*. 2021;77:539–46.
13. Zwarenstein M, Treweek S, Loudon K. PRECIS-2 helps researchers design more applicable RCTs while CONSORT Extension for pragmatic trials helps knowledge users decide whether to apply them. *J Clin Epidemiol*. 2017;84:27–9.
14. Patsopoulos NA. A pragmatic view on pragmatic trials. *Dialogues Clin Neurosci*. 2011;13:217–24.
15. Kent DM, Kitsios G. Against pragmatism: on efficacy, effectiveness and the real world. *Trials*. 2009;10:48.
16. Ernst E, Canter PH. Limitations of pragmatic trials. *Postgrad Med J*. 2005;81:203–203.
17. Alvarez G, Solà I, Sitjà-Rabert M, Fort-Vanmeerhaeghe A, Gich I, Fernández C, et al. A methodological review revealed that reporting of trials in manual therapy has not improved over time. *J Clin Epidemiol*. 2020;121:32–44.
18. Oliveira CB, Maher CG, Pinto RZ, Traeger AC, Lin C-WC, Chenot J-F, et al. Clinical practice guidelines for the management of non-specific low back pain in primary care: an updated overview. *Eur Spine J*. 2018;27:2791–803.
19. Qaseem A, Wilt TJ, McLean RM, Forciea MA, for the Clinical Guidelines Committee of the American College of Physicians. Noninvasive treatments for Acute, Subacute, and chronic low back Pain: a clinical practice Guideline from the American College of Physicians. *Ann Intern Med*. 2017;166:514.
20. Ninot G. Evaluation of non-pharmacological interventions. *Non-Pharmacological Interventions*. 2021;221–38.
21. Dieppe P. Complex interventions. *Musculoskeletal Care*. 2004;2:180–6.
22. Richards DA, Hallberg JR. Complex interventions in Health: an overview of research methods. Routledge; 2015.
23. Walach H, Falkenberg T, Fønnebo V, Lewith G, Jonas WB. Circular instead of hierarchical: methodological principles for the evaluation of complex interventions. *BMC Med Res Methodol*. 2006;6:29.
24. Minary L, Trompette J, Kivits J, Cambon L, Tarquinio C, Alla F. Which design to evaluate complex interventions? Toward a methodological framework through a systematic review. *BMC Med Res Methodol*. 2019;19:1–9.
25. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, et al. Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ*. 2008;337:a1655.
26. Pinnock H, Epiphaniou E, Sheikh A, Griffiths C, Eldridge S, Craig P, et al. Developing standards for reporting implementation studies of complex interventions (StaRI): a systematic review and e-Delphi. *Implement Sci*. 2015;10:42.
27. Bishop MD, Torres-Cueco R, Gay CW, Lluch-Girbés E, Beneciuk JM, Bialosky JE. What effect can manual therapy have on a patient's pain experience? *Pain Manag*. 2015;5:455–64.
28. Daniel Maddox C, Subiela JA, Young JL, Rhon DI. TITLE: over half of clinical trials of mobilization and manipulation for patients with low back Pain May have Limited Real-World Applicability. *A Systematic Review of 132 Clinical Trials*.
29. Taljaard M, Nicholls SG, Howie AH, Nix HP, Carroll K, Moon PM, et al. An analysis of published trials found that current use of pragmatic trial labels is uninformative. *J Clin Epidemiol*. 2022.
30. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Rev Esp Cardiol*. 2021;74:790–9.
31. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. *Syst Rev*. 2016;5:210.
32. Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. *BMJ*. 2015;350:h2147.
33. Zhang Y-Q, Jiao R-M, Witt CM, Lao L, Liu J-P, Thabane L, et al. How to design high quality acupuncture trials—a consensus informed by evidence. *BMJ*. 2022;376.
34. Alvarez G, Núñez-Cortés R, Solà I, Sitjà-Rabert M, Fort-Vanmeerhaeghe A, Fernández C, et al. Sample size, study length, and inadequate controls were the most common self-acknowledged limitations in manual therapy trials: a methodological review. *J Clin Epidemiol*. 2021;130:96–106.
35. Devos F, Foissac F, Bouazza N, Ancel P-Y, Tréluyer J-M, Chappuy H. Study characteristics impacted the pragmatism of randomized controlled trial published in nursing: a meta-epidemiological study. *J Clin Epidemiol*. 2019;116:18–25.
36. Elder WG, Munk N. Using the pragmatic-explanatory Continuum Indicator Summary (PRECIS) model in Clinical Research: application to refine a practice-based Research Network (PBRN) Study. *J Am Board Family Med*. 2014;27:846–54.
37. Rosas LG, Lv N, Azar K, Xiao L, Yank V, Ma J. Applying the pragmatic–explanatory Continuum Indicator Summary Model in a primary care–based lifestyle intervention trial. *Am J Prev Med*. 2015;49(3):S208–14.
38. Lee J, Cho J-H, Kim K-W, Lee J-H, Kim M-R, Kim J, et al. Chuna Manual Therapy vs Usual Care for patients with nonspecific chronic Neck Pain: a Randomized Clinical Trial. *JAMA Netw Open*. 2021;4:e2113757.
39. Nguyen AP, Pitance L, Mahaudens P, Detrembleur C, David Y, Hall T, et al. Effects of Mulligan Mobilization with Movement in Subacute lateral ankle sprains: a pragmatic Randomized Trial. *J Man Manip Ther*. 2021;29:341–52.
40. Groisman S, Malysz T, de Souza da Silva L, Rocha Ribeiro Sanches T, Camargo Bragante K, Locatelli F, et al. Osteopathic manipulative treatment combined with exercise improves pain and disability in individuals with non-specific chronic neck pain: a pragmatic randomized controlled trial. *J Bodyw Mov Ther*. 2019;24:189–95.
41. Wilkey A, Gregory M, Byfield D, McCarthy PW. A comparison between chiropractic management and pain clinic management for chronic low-back pain in a national health service outpatient clinic. *J Altern Complement Med*. 2008;14:465–73.
42. Finch P, Bessonnette S. A pragmatic investigation into the effects of massage therapy on the self-efficacy of multiple sclerosis clients. *J Bodyw Mov Ther*. 2014;18:11–6.
43. Griswold D, Learnan K, O'Halloran B, Cleland J. A preliminary study comparing the use of cervical/upper thoracic mobilization and manipulation for individuals with mechanical neck pain. *J Man Manip Ther*. 2015;23:75–83.
44. Cross J, Elender F, Barton G, Clark A, Shepstone L, Blyth A, et al. A randomised controlled equivalence trial to determine the effectiveness and cost-utility of manual chest physiotherapy techniques in the management of exacerbations of chronic obstructive pulmonary disease (MATREX). *Health Technol Assess*. 2010;14.
45. Poole H, Glenn S, Murphy P. A randomised controlled study of reflexology for the management of chronic low back pain. *Eur J Pain*. 2007;11:878–87.
46. Sharp DM, Walker MB, Chaturvedi A, Upadhyay S, Hamid A, Walker AA, et al. A randomised, controlled trial of the psychological effects of reflexology in early breast cancer. *Eur J Cancer*. 2010;46:312–22.
47. Stockkendahl MJ, Christensen HW, Vach W, Hoiland-Carlson PF, Hagfelt T, Hartvigsen J. A randomized clinical trial of chiropractic treatment and self-management in patients with acute musculoskeletal chest pain: 1-year follow-up. *J Manipulative Physiol Ther*. 2012;35:254–62.
48. Goertz CM, Long CR, Hondras MA, Petri R, Delgado R, Lawrence DJ, et al. Adding chiropractic manipulative therapy to standard medical care for patients with acute low back pain: results of a pragmatic randomized comparative effectiveness study. *Spine*. 2013;38:627–34.
49. Attias S, Sivan K, Avneri O, Sagee A, Ben-Arye E, Grinberg O, et al. Analgesic effects of Reflexology in patients Undergoing Surgical procedures: a Randomized Controlled Trial. *J Altern Complement Med*. 2018;24:809–15.
50. Park S-Y, Hwang E-H, Cho J-H, Kim K-W, Ha I-H, Kim M-R, et al. Comparative effectiveness of Chuna Manipulative Therapy for Non-acute Lower Back Pain: a Multi-center, pragmatic, randomized controlled trial. *J Clin Med Res*. 2020;9.
51. Hay EM, Mullis R, Lewis M, Vohora K, Main CJ, Watson P, et al. Comparison of physical treatments versus a brief pain-management programme for back pain in primary care: a randomised clinical trial in physiotherapy practice. *Lancet*. 2005;365:2024–30.
52. Dissing KB, Hartvigsen J, Wedderkopp N, Hestbæk L. Conservative care with or without manipulative therapy in the management of back and/or neck pain in Danish children aged 9–15: a randomised controlled trial nested in a school-based cohort. *BMJ Open*. 2018;8:e021358.
53. Wyatt K, Edwards V, Franck L, Britten N, Creanor S, Maddick A, et al. Cranial osteopathy for children with cerebral palsy: a randomised controlled trial. *Arch Dis Child*. 2011;96:505–12.

54. Goertz CM, Long CR, Vining RD, Pohlman KA, Walter J, Coulter I. Effect of Usual Medical Care Plus Chiropractic Care vs Usual Medical Care alone on Pain and Disability among US Service Members with Low Back Pain: a comparative effectiveness clinical trial. *JAMA Netw Open*. 2018;1:e180105.
55. Castien RF, van der Windt DAWM, Grooten A, Dekker J. Effectiveness of manual therapy for chronic tension-type headache: a pragmatic, randomised, clinical trial. *Cephalalgia*. 2011;31:133–43.
56. Dzedzic K, Hill J, Lewis M, Sim J, Daniels J, Hay EM. Effectiveness of manual therapy or pulsed shortwave diathermy in addition to advice and exercise for neck disorders: a pragmatic randomized controlled trial in physical therapy clinics. *Arthr Rheum*. 2005;53:214–22.
57. Mafetoni RR, Shimo AKK. Effects of acupressure on progress of labor and cesarean section rate: randomized clinical trial. *Rev Saude Publica*. 2015;49:9.
58. Miller JE, Newell D, Bolton JE. Efficacy of chiropractic manual therapy on infant colic: a pragmatic single-blind, randomized controlled trial. *J Manipulative Physiol Ther*. 2012;35:600–7.
59. Walach H, G  thlin C, K  nig M. Efficacy of massage therapy in chronic pain: a pragmatic randomized trial. *J Altern Complement Med*. 2003;9:837–46.
60. Harper B, Steinbeck L, Aron A. Fascial manipulation vs. standard physical therapy practice for low back pain diagnoses: a pragmatic study. *J Bodyw Mov Ther*. 2019;23:115–21.
61. Bergman GJ, Winters JC, Groenier KH, Meyboom-de Jong B, Postema K, van der Heijden GJ. Manipulative therapy in addition to usual care for patients with shoulder complaints: results of physical examination outcomes in a randomized controlled trial. *J Manipulative Physiol Ther*. 2010;33:96–101.
62. Groeneweg R, van Assen L, Kropman H, Leopold H, Mulder J, Smits-Engelsman BCM, et al. Manual therapy compared with physical therapy in patients with non-specific neck pain: a randomized controlled trial. *Chiropr Man Th*. 2017;25:12.
63. Hoving M. Therapy. Physical therapy, or continued care by the General Practitioner for patients with Neck Pain Long-Term results from a pragmatic Randomized Clinical Trial. 2006.
64. Lilje S, Friberg H, Wykman A, Skillgate E. Naprapathic manual therapy or conventional orthopedic care for outpatients on orthopedic waiting lists? A pragmatic randomized controlled trial. *Clin J Pain*. 2010;26:602–10.
65. Schwela F, Rother K, Rother D, R  tz M, Resch K-L. Osteopathic Manipulative Therapy in Women with Postpartum Low Back Pain and disability: a pragmatic randomized controlled trial. *J Am Osteopath Assoc*. 2015;115:416–25.
66. Williams NH, Wilkinson C, Russell I, Edwards RT, Hibbs R, Linck P, et al. Randomized osteopathic manipulation study (ROMANS): pragmatic trial for spinal pain in primary care. *Fam Pract*. 2003;20:662–9.
67. Gemmell H, Miller P. Relative effectiveness and adverse effects of cervical manipulation, mobilisation and the activator instrument in patients with sub-acute non-specific neck pain: results from a stopped randomised trial. *Chiropr Osteopathy*. 2010;18.
68. Castro-S  nchez AM, Lara-Palomo IC, Matar  n-Pe  narrocha GA, Fern  ndez-de-las-Pe  as C, Saavedra-Hern  ndez M, Cleland J, et al. Short-term effectiveness of spinal manipulative therapy versus functional technique in patients with chronic nonspecific low back pain: a pragmatic randomized controlled trial. *Spine J*. 2016;16:302–12.
69. Bronfort G, H  ndras MA, Schulz CA, Evans RL, Long CR, Grimm R. Spinal manipulation and home exercise with advice for subacute and chronic back-related leg pain: a trial with adaptive allocation. *Ann Intern Med*. 2014;161:381–91.
70. Evans R, Haas M, Schulz C, Leininger B, Hanson L, Bronfort G. Spinal manipulation and exercise for low back pain in adolescents: a randomized trial. *Pain*. 2018;159:1297–307.
71. Georgoudis G, Felah B, Nikolaidis P, Damigos D. The effect of myofascial release and microwave diathermy combined with acupuncture versus acupuncture therapy in tension-type headache patients: a pragmatic randomized controlled trial. *Physiother Res Int*. 2018;23:e1700.
72. Skillgate E, Bohman T, Holm LW, Ving  rd E, Alfredsson L. The long-term effects of naprapathic manual therapy on back and neck pain - results from a pragmatic randomized controlled trial. *BMC Musculoskelet Disord*. 2010;11:26.
73. Eklund A, Jensen I, Lohela-Karlsson M, Hagberg J, Leboeuf-Yde C, Kongsted A, et al. The Nordic Maintenance Care program: effectiveness of chiropractic maintenance care versus symptom-guided treatment for recurrent and persistent low back pain—A pragmatic randomized controlled trial. *PLoS ONE*. 2018;13:e0203029.
74. Lim K-T, Hwang E-H, Cho J-H, Jung J-Y, Kim K-W, Ha I-H, et al. Comparative effectiveness of Chuna manual therapy versus conventional usual care for non-acute low back pain: a pilot randomized controlled trial. *Trials*. 2019;20:216.
75. Kim B-J, Park A-L, Hwang M-S, Heo I, Park S-Y, Cho J-H et al. Comparative effectiveness and Safety of Concomitant Treatment with Chuna Manual Therapy and Usual Care for Whiplash injuries: a Multicenter Randomized Controlled Trial. *Int J Environ Res Public Health*. 2022;19.
76. Lynen A, Sch  mitz M, Vahle M, J  kel A, R  tz M, Schwela F. Osteopathic treatment in addition to standard care in patients with gastroesophageal reflux disease (GERD) - a pragmatic randomized controlled trial. *J Bodyw Mov Ther*. 2022;29:223–31.
77. Alvarez G, Cerritelli F, Urrutia G. Using the template for intervention description and replication (TIDieR) as a tool for improving the design and reporting of manual therapy interventions. *Man Ther*. 2016;24:85–9.
78. Cashin AG, Lee H, Bagg MK, O'Hagan E, Traeger AC, Kamper SJ, et al. A systematic review highlights the need to improve the quality and applicability of trials of physical therapy interventions for low back pain. *J Clin Epidemiol*. 2020;126:106–15.
79. Kamper SJ, Moseley AM, Herbert RD, Maher CG, Elkins MR, Sherrington C. 15 years of tracking physiotherapy evidence on PEDro, where are we now? *Br J Sports Med*. 2015;49:907–9.
80. Palakshappa JA, Gibbs KW, Lannan MT, Cranford AR, Taylor SP. Systematic review of the pragmatism of pragmatic critical care trials. *Crit Care Explorations*. 2022;4:e0738.
81. Sepehrvand N, Alemayehu W, Das D, Gupta AK, Gouda P, Ghimire A, et al. Trends in the explanatory or pragmatic nature of Cardiovascular clinical trials over 2 decades. *JAMA Cardiol*. 2019;4:1122–8.
82. Dal-R   R, de Boer A, James SK. The design can limit PRECIS-2 retrospective assessment of the clinical trial explanatory/pragmatic features. *J Clin Epidemiol*. 2020;126:193–201.
83. Averitt AJ, Ryan PB, Weng C, Perotte A. A conceptual framework for external validity. *J Biomed Inf*. 2021;121:103870.
84. Dal-R   R. Pragmatic trials, blinding, placebos, and the usefulness of the PRECIS-2 tool. *Eur J Clin Pharmacol*. 2021;77:1071–2.
85. Zwarenstein M, Thorpe K, Treweek S, Loudon K. PRECIS-2 for retrospective assessment of RCTs in systematic reviews. *J Clin Epidemiol*. 2020;126:202–6.
86. Rossetini G, Camerone EM, Carlino E, Benedetti F, Testa M. Context matters: the psychoneurobiological determinants of placebo, nocebo and context-related effects in physiotherapy. *Arch Physiother*. 2020;10:11.
87. Rossetini G, Carlino E, Testa M. Clinical relevance of contextual factors as triggers of placebo and nocebo effects in musculoskeletal pain. *BMC Musculoskelet Disord*. 2018;19:27.
88. Bialosky JE, Bishop MD, Penza CW. Placebo mechanisms of Manual Therapy: A Sheep in Wolf's Clothing? *J Orthop Sports Phys Ther*. 2017;47:301–4.
89. Bialosky JE, Beneciuk JM, Bishop MD, Coronado RA, Penza CW, Simon CB, et al. Unraveling the mechanisms of Manual Therapy: modeling an Approach. *J Orthop Sports Phys Ther*. 2018;48:8–18.
90. Zwarenstein M, Howie A. Blinding, pragmatism, and the PRECIS-2 tool for designing and assessing randomized trials. *Eur J Clin Pharmacol*. 2021;77:1069–70.
91. Leviton LC. Generalizing about Public Health Interventions: a mixed-methods Approach to External Validity. *Annu Rev Public Health*. 2017;38:371–91.
92. Hohenschurz-Schmidt DJ, Cherkin D, Rice ASC, Dworkin RH, Turk DC, McDermott MP et al. Research objectives and general considerations for pragmatic clinical trials of pain treatments: IMMPACT statement. *Pain*. 2023.
93. Zhang YQGG. How to design high quality acupuncture trials—a consensus informed by evidence. *BMJ*. 2022;377:o1046.
94. Neta G, Johnson KE. Informing real-world practice with real-world evidence: the value of PRECIS-2. *BMC Med*. 2018;16:76.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

4.2 Resultados del segundo trabajo

4.2.1 Evaluación del reporte a través del cumplimiento de los ítems de la guía CONSORT

El cuarenta y uno por ciento (16/39) de los estudios declararon estar reportados según la guía CONSORT. Aun así, sólo un estudio [130] informó que sus hallazgos se adhirieron a la extensión CONSORT para estudios pragmáticos. Uno de cada tres estudios de la muestra cumplió con más del 70% de los ítems CONSORT (13/39), entre el 50% y el 70% en el 49% de los estudios (19/39) y menos del 50% de los ítems CONSORT en 13% de la muestra (5/39). Esta adherencia varió entre los ítems, como se puede observar en la Tabla 3 (de la publicación respectiva). Si bien algunos ítems se reportaron algo mejor en el período post-CONSORT, esta mejora no afectó al resultado final obtenido. Los ítems reportados por todos los estudios incluidos fueron, un resumen estructurado (1b), criterios de elegibilidad claros (4a) y métodos estadísticos (12a). Con respecto a la extensión CONSORT para estudios pragmáticos, todos los ítems se informaron en más del 72% (28/39) de los casos, excepto el ítem 5d (recursos adicionales agregados a los entornos habituales para implementar la intervención). Aunque la descripción de la intervención para cada grupo de estudio (ítem 5a) fue buena en más del 70% (28/39) de los estudios, los autores no informaron suficientemente los detalles de la intervención, como los diferentes componentes de la intervención, la estandarización o individualización de la intervención y los recursos adicionales para parecerse a la práctica clínica. Además, menos del 20% (7/39) de los estudios informaron si se evaluó la adherencia de los participantes o los terapeutas y cómo. Tampoco se informó sobre los "efectos adversos" y "número de registro del estudio y dónde se puede evaluar el protocolo" en casi el 50% (16/39) de la muestra. El treinta por ciento (12/39) de los estudios no informó estrategias para limitar el sesgo cuando no era posible el enmascaramiento, mecanismos para implementar la asignación de la secuencia aleatoria, la forma en que se calculó el tamaño de la muestra o el entorno donde se recopilaron los datos. El veinticinco por ciento (10/39) de los estudios no informaron

70

sobre la generalización de los resultados del estudio ni explicaron por qué no fue posible el enmascaramiento.

4.2.2 Evaluación del riesgo de sesgo

La valoración general del riesgo de sesgo se resume en la Figura 3 (de la publicación respectiva). El sesgo de realización y de detección fueron los dominios con mayor riesgo de sesgo (puntuaron alto en el 90%, 35/39 y el 77%; 30/39 de los estudios incluidos, respectivamente). El sesgo de reporte no estaba claro en casi la mitad de la muestra (46%; 18/39).

4.2.3 Relación entre el pragmatismo, la calidad del reporte y el riesgo de sesgo

Los resultados sobre la actitud pragmática de los EPs en TM revelaron que el 61% (24/39) de los estudios eran moderadamente pragmáticos, el 36% (14/39) muy pragmáticos y el 3% (1/37) ligeramente pragmáticos. No se encontró relación entre la actitud pragmática y el buen reporte de los estudios, excepto en el ítem 25 del CONSORT (Fuentes de financiación y otros apoyos) ($p=0,005$). Asimismo, no se encontró relación entre el RoB y la actitud pragmática de los estudios.

La comparación de los estudios con un RoB bajo versus alto/poco claro según el cumplimiento de los ítems de la guía CONSORT mostraron que el porcentaje de cumplimiento a CONSORT fue mayor en los estudios con un RoB bajo en los siguientes dominios: Generación de la secuencia aleatoria, ocultamiento de la asignación, datos incompletos y reporte selectivo.

4.2.4 Segunda publicación

Roura, Alvarez, Hohenschurz-Schmidt, D., Solà, Núñez-Cortés, R., Bracchiglione, Fernández-Jané, C., Phalip, Gich, Sitjà-Rabert, M., & Urrútia. (2025). **A call for improving the Internal validity and the reporting of manual therapy trials self-labelled as pragmatic: A methodological review.** *International Journal of Osteopathic Medicine: IJOM*, 100754, 100754.



A call for improving the internal validity and the reporting of manual therapy trials self-labelled as pragmatic: A methodological review

S. Roura^a, G. Alvarez^{b,c,*}, D. Hohenschurz-Schmidt^{d,e}, I. Solà^{b,f,g}, R. Núñez-Cortés^h, J. Bracchiglione^{b,f,i}, C. Fernández-Jané^{c,j,k}, J. Phalip^{l,m}, I. Gich^{f,n}, M. Sitjà-Rabert^{c,j}, G. Urrútia^{b,f,g}

^a PhD Student in Biomedical Research Methodology and Public Health in the Medical Department of the Universitat Autònoma de Barcelona, Barcelona, Spain

^b Iberoamerican Cochrane Centre – Biomedical Research Institute Sant Pau, IIB Sant Pau, Barcelona, Spain

^c Department of Physical Therapy, Faculty of Health Science Blanquerna, Ramon Llull University, Barcelona, Spain

^d Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, London, United Kingdom

^e UCO School of Osteopathy, Health Sciences University, London, United Kingdom

^f Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP), Madrid, Spain

^g Universitat Autònoma de Barcelona, Bellaterra, Spain

^h Department of Physical Therapy, Faculty of Medicine, University of Chile, Santiago, Chile

ⁱ Interdisciplinary Centre for Health Studies (CIESAL), Universidad de Valparaíso, Viña Del Mar, Chile

^j Global Research on Wellbeing (GROW) Research Group, Ramon Llull University, Barcelona, Spain

^k Tecnocampus, Universitat Pompeu Fabra, Marató-Maresme, Barcelona, Spain

^l Institut ANALGESIA, Faculté de Médecine, Université Clermont Auvergne, Clermont-Ferrand, France

^m CHU Clermont-Ferrand, Inserm 1107 Neuro-Dol, Service de Pharmacologie Médicale, Université Clermont Auvergne, Clermont-Ferrand, France

ⁿ Institut d'Investigació Biomèdica Sant Pau (IIB SANT PAU), Barcelona, Spain

ARTICLE INFO

Keywords:

Pragmatic clinical trials
Musculoskeletal manipulations
Manual therapy
Generalizability
CONSORT
Risk of bias

ABSTRACT

Objectives: This study provides further data from a review assessing methodological characteristics of pragmatic randomised controlled trials (pRCTs) published in manual therapy (MT). In this second part, we aim to assess the report, the risk of bias (RoB), and the relationship between these items and the pragmatism scores of the self-labelled pRCTs in the MT field.

Study design and setting: We searched MEDLINE and the Cochrane Central Register of Controlled Trials for pRCTs in the MT field from inception to January 2024. Two independent reviewers screened the trials using several CONSORT extensions and assessed them using the Cochrane Risk of Bias tool. We performed a descriptive analysis using frequencies and percentages and a relation analysis between the trials' pragmatism, their reporting, and their RoB.

Results: We included 39 self-labelled MT pRCTs. Compliance with CONSORT items was higher than 70 % in one-third of the included trials (13/39) but varied across items. Performance and detection bias were the main threats to internal validity (we rated 90 %, 35/39, and 77 %; 30/39 of trials at high risk of bias, respectively). Selective reporting bias was unclear in almost half of the sample (46 %; 18/39). No relation was found between the highly pragmatic attitude and good reporting except for CONSORT item 25 (Sources of funding and other support) (p 0.006). No relation was found between the RoB and the pragmatic attitude of the studies. The percentage of compliance with CONSORT items was higher in the trials with low RoB.

Conclusion: Pragmatic trials in MT have significant methodological limitations, and their reporting is suboptimal. Nonetheless, trials with less risk of bias had higher compliance with CONSORT items.

Implications for practice

- Adherence to the CONSORT checklist is suboptimal in MT pRCTs

- Performance and detection biases are always penalised in MT pRCTs because of the intrinsic difficulties MT interventions present on

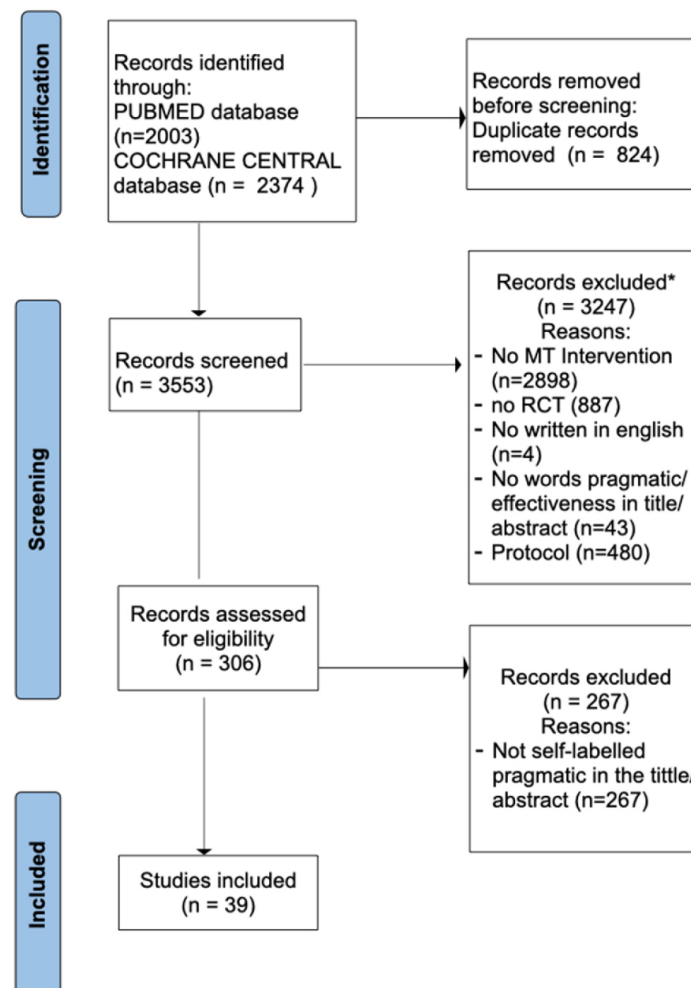
* Corresponding author. Iberoamerican Cochrane Center, C. Sant Antoni Maria Claret 167, Pavelló 18, Planta 0, 08025 Barcelona, Catalunya, Spain.
E-mail address: gerardalv@gmail.com (G. Alvarez).

<https://doi.org/10.1016/j.ijosm.2025.100754>

Received 12 November 2024; Received in revised form 5 February 2025; Accepted 5 February 2025

Available online 8 February 2025

1746-0689/© 2025 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.



* It should be noted that some records have more than one reason for being excluded.

Fig. 1. PRISMA flowchart (Reproduced with permission from Roura et al., 2024 [17]).

blinding. Moreover, there is a concern about the selective reporting of those trials.

- Good reporting of some CONSORT items directly relates to a low RoB in MT pRCTs.
- Adherence to reporting guidelines is imperative.

1. Introduction

Despite the recent surge in manual therapy (MT) research, numerous methodological flaws continue to be reported by authors in this field [1–7]. Achieving a balance between external and internal validity in trials is fundamental to all research fields [8]. This challenge in balancing external and internal validity has already been observed in RCTs evaluating Complex Interventions (CI) [9–12], particularly in the

MT field [13,14]. Among other strategies, some have proposed a deeper focus on the pragmatic attitude of randomised controlled trials (RCTs) to overcome methodological deficiencies [9,10]. It should be noted that interest in pragmatic randomised controlled trials (pRCTs) has grown globally over the last few years [15,16]. Despite this, it remains unclear how this research design has been implemented in MT research. Previous findings from the current methodological review on the level of pragmatism in self-labelled pRCTs demonstrate that MT trials exhibit a moderately pragmatic approach (mean score of 3.5 on the PRECIS-2 tool) [17].

On the flip side, studies have consistently shown a strong correlation between the overall methodological quality of MT literature studies and the quality of their reporting [3,6,7,18,19]. Additionally, reporting plays a crucial role in pRCTs, serving both to justify an author's

Table 1
Identity elements of the included studies.

STUDIES INCLUDED	Publication Year	Number of participants	Condition treated	Intervention type	PRECIS-2 SCORE
Lee et al. [30]	2021	108	Nonspecific Chronic Neck Pain	Chuna manual therapy	3.4
Nguyen et al. [31]	2021	61	Lateral Ankle Sprains	Mulligan	2.7
Groisman et al. [32]	2019	90	Non-specific chronic neck pain	Osteopathic treatment	3.9
Wilkey A et al. [33]	2008	30	Chronic low-back pain	Chiropractic treatment	2.9
Finch P et al. [34]	2014	15	Multiple sclerosis	Massage therapy	3.9
Griswold et al. [35]	2015	22	Mechanical neck pain	Spinal manipulation	3.1
Cross et al. [36]	2010	527	Chronic obstructive pulmonary disease	Manual chest physiotherapy	4.1
Poole et al. [37]	2007	243	Chronic low back pain	Reflexology	3.2
Sharp et al. [38]	2010	183	Breast cancer	Reflexology	4
Stochkendahl et al. [39]	2012	115	Acute musculoskeletal chest pain	Chiropractic manipulative therapy	3.5
Goertz et al. [40]	2013	91	Acute Low Back Pain	Chiropractic manipulative therapy	3.1
Attias et al. [41]	2018	164	Surgical Procedures	Reflexology	3.6
Park et al. [42]	2020	194	Non-Acute Lower Back Pain	Chuna manipulative therapy	3.8
Hay et al. [43]	2005	402	Back pain	Manual therapy	4
Dissing et al. [44]	2018	238	Back and/or neck pain	Chiropractic manipulative therapy	3.6
Wyatt et al. [45]	2011	142	Cerebral palsy	Cranial osteopathy	3.8
Goertz et al. [46]	2018	750	Low Back Pain	Chiropractic treatment	4.3
Castien et al. [47]	2011	82	Chronic tension-type headache	Manual therapy	3.7
Dziedziec et al. [48]	2005	350	Neck Disorders	Manual therapy	4.5
Mafetoni et al. [49]	2015	156	Labor and cesarean	Acupressure	3.7
Miller et al. [50]	2012	104	Infant colic	Chiropractic manipulative therapy	3.3
Walach et al. [51]	2003	29	Chronic pain	Massage	4.4
Harper et al. [52]	2019	102	Low back pain	Fascial manipulation	3.9
Bergman et al. [53]	2010	150	Shoulder complaints	Manipulative therapy	4.2
Greeneweg et al. [54]	2017	181	Non-specific neck pain	Manual therapy	3.9
Hoving et al. [55]	2006	183	Neck Pain	Manual therapy/Physical therapy	4.3
Lilje et al. [56]	2010	78	Outpatients on orthopedic waiting lists	Naprapathic manual therapy	3.9
Schwerla et al. [57]	2015	80	Postpartum Low Back Pain	Osteopathy	3.9
Williams et al. [58]	2003	201	Spinal pain	Spinal manipulation	3.7
Gemmell et al. [59]	2010	47	Sub-acute non-specific neck pain	Spinal manipulation	3.1
Gastro-Sánchez et al. [60]	2016	62	Chronic nonspecific low back pain	Spinal manipulation/functional technique	3.3
Evans et al. [62]	2018	185	Low back pain	Spinal manipulation	3.3
Bronfort et al. [61]	2014	192	Chronic back-related leg pain	Spinal manipulation	3.3
Georgondis et al. [63]	2017	44	Tension-type headache	Myofascial release	2.4
Skillgate et al. [64]	2010	409	Back and neck pain	Naprapathy	3.2
Eklund et al. [65]	2018	328	Recurrent and persistent low back pain	Chiropractic care	4.8
Lim et al. [66]	2016	60	Non-acute low back pain	Chuna manual therapy	3.1
Kim et al. [67]	2022	70	Whiplash Injuries	Osteopathic treatment	4.6
Lynen et al. [68]	2022	132	Gastroesophageal Reflux Disease	Chuna manual therapy	3.4

pragmatic approach and provide context for the trial's pragmatic attitude [15,20]. To enhance the reporting of pragmatic trials and focus on applicability, an extension of the CONSORT statement - the Consolidated Standards of Reporting Trials - has been developed [21]. Several studies have examined adherence to the CONSORT reporting guideline (RG) in rehabilitation RCTs [1,6]. Previous research has shown that high-impact rehabilitation journals do not mention RG use, and a high percentage of those who claimed to use an RG did not do so appropriately [22]. Specifically, in MT RCTs, it was concluded that the quality of reporting has not improved over time. The quality of reporting can directly influence the risk of bias assessment, which is crucial for estimating the treatment effect. In this regard, a recent review by Arienti et al. concluded that bias can influence the treatment effect in different directions [7].

To our knowledge, there has been no systematic evaluation of compliance with CONSORT items or the relationship between completeness of reporting and RoB of pRCTs.

We reported the pragmatic attitude of self-labelled pragmatic trials published in the MT field elsewhere [17]. The primary aim of this second paper is to assess the quality of reporting and internal validity of MT pRCTs. As a secondary objective, we examined the relation between the reporting and the risk of bias in conjunction with the pragmatism of the trials.

2. Methods

2.1. Protocol registration

We conducted a review to describe and assess the methodological characteristics of pragmatic trials on MT interventions. We reported the review according to the PRISMA guidelines [23] (Supplementary file 1). We prospectively registered the protocol in the Open Science Framework (DOI 10.17605/OSF.IO/WKEPZ).

2.2. Eligibility criteria

We included RCT reports with the words 'effectiveness' (related to an intervention), 'pragmatic' or 'naturalistic' (related to the methodological design) in either the title or abstract [16,24]. The eligible resources must comprise either one or a combination of manual techniques, including soft tissue techniques, joint mobilisations or manipulations, massage, myofascial release, nerve manipulation, strain/counterstrain, and acupressure. No limitations have been imposed with regard to population, comparators, or outcome measures. Experimental interventions delivered through electrotherapy, kinesiotaping, dry needling, acupuncture, drugs, active exercises or a combination of therapies not involving MT intervention were excluded from the study. Additionally, articles not written in English, protocols and

Table 2
Main characteristics of pRCTs in MT.

Number of participants (Mean)	169,23 (SD = 150,7)
N of participants	Per cent (n ^a)
<50	15 % (6)
51–100	23 % (9)
101–200	38 % (15)
>201	23 % (9)
EXPERIMENTAL INTERVENTION	
Combination of non-protocolised techniques	54 % (21)
Isolated non-protocolised technique	15 % (6)
Protocol of a combination of techniques	13 % (5)
Combination of non-protocolised therapies	10 % (4)
Protocol of an isolated technique	5 % (2)
Protocol of a combination of therapies	2 % (1)
CONTROL INTERVENTION	
2 arms	
test treatment vs other active intervention	31 % (12)
test treatment vs placebo	2 % (1)
test treatment vs usual care	33 % (13)
test treatment vs no intervention	13 % (5)
3 arms	
test treatment vs 2 other active interventions	5 % (2)
test treatment vs 1 other intervention and 1 placebo	2 % (1)
test treatment vs 1 intervention and 1 usual care	10 % (4)
test treatment vs 1 other intervention and 1 no intervention	2 % (1)
BLINDING (YES)	
Participants	8 % (3)
Therapists	0 % (0)
Outcome assessors	69 % (27)
Statistician	43 % (17)
SETTING	
Multicentric	56 % (22)
Unicentric	33 % (13)
Unclear report	10 % (4)
FOLLOW-UP	
No follow-up	18 % (7)
<2 weeks	0 % (0)
2–4 weeks	2 % (1)
4–12 weeks	10 % (4)
3–6 months	18 % (7)
6–12 months	20 % (8)
>1 year	26 % (10)
Individualised ^a	5 % (2)

^a Two studies used an individualised follow-up (trials assessing how many weeks patients remained pain-free).

^b N of studies = 39.

poster/conference presentations were also excluded.

2.3. Search strategy

We conducted an extensive search of MEDLINE and the Cochrane Central Register of Controlled Trials from inception to January 2024. Our search strategy integrated controlled vocabulary with search terms pertinent to the field of MT and the design of interest (Supplementary file 2). The search strategy was designed by an experienced methodologist who is a member of our research team (IS).

2.4. Study selection

We imported the records into Rayyan software (www.rayyan.org) [25]. After removing duplicate entries, two reviewers (SR, GA) screened the references independently based on their titles and abstracts and resolved disagreements through discussion. For this paper, only the effectiveness studies self-described as pragmatic in the title and/or abstract were included.

2.5. Data collection process

Two independent reviewers (SR and one additional reviewer from GA, RN, JB, DH, CF and JP) extracted data from the included studies and resolved any disagreements by consensus. The extraction data form was designed by three authors (SR, GA, GU) and was used to collect data. After piloting five studies, the data extraction form was revised to incorporate suggestions from all reviewers to improve consistency. The team was also provided with a guideline for reviewers (Supplementary file 3).

2.6. Data items

The data extraction included the following elements.

1. Quality of report: all items from the CONSORT reporting guideline and the extension for pragmatic trials and non-pharmacological interventions were included [26–28].
2. Risk of Bias: RoB Cochrane tool [29].

Other items included in the data extraction have been assessed and reported elsewhere [17]. They included the bibliometric identification elements, the intent of the trial, the rationale of the intervention given by the authors, the experimental and control interventions and the pragmatism of the trials with the PRECIS-2 tool (Supplementary file 3).

2.7. Data analysis

A descriptive analysis was undertaken on the categorical variables, and the findings were displayed as absolute and relative frequencies. The quantitative variables were described in terms of mean and standard deviation. Ordinal variables were defined by median and interquartile range. Chi-square was employed for categorical variables, and Kruskal-Wallis was used for ordinal variables to compare variables. We conducted a post-hoc analysis using Mann-Whitney test for Kruskal-Wallis. We compared in a post-hoc analysis the CONSORT results between two periods (before 2009 and after 2011). Additionally, we used the Spearman test to examine the correlation between CONSORT and RoB tools and the pragmatic attitude. Findings regarding the pragmatic attitude have been recently published elsewhere [17]. We explored the link between these results and the reporting and RoB of pRCTs. The significance level was set at 5 % ($\alpha = 0.05$). All data were analysed with IBM-SPSS (V26.0) software.

3. RESULTS

3.1. Study selection

After the removal of duplicate references, the search identified 3553 articles. After screening the title, abstract, and full-text disponibility, the final sample comprised 306 studies. Among these, 39 were identified as self-labelled pragmatic in the title and/or abstract, whereas 267 were 'effectiveness' Randomised Controlled Trials (RCTs) not self-labelled as pragmatic. Subsequently, for the purpose of this study, 39 self-identified pragmatic randomised controlled trials were included after screening [30–68]. Fig. 1 provides the PRISMA diagram, and the identity elements of the included studies are presented in Table 1.

3.2. Characteristics and pragmatism of trials

The main characteristics of our sample and the pragmatism of the MT pRCTs are summarised in Table 2 and reported extensively elsewhere [17]. The mean PRECIS-2 score was 3.5. Collectively, they showed a moderately pragmatic attitude. The wheel PRECIS-2 tool diagram is represented in Fig. 2. A summary of the PRECIS-2, CONSORT and RoB assessment for each included paper can be found in supplementary file

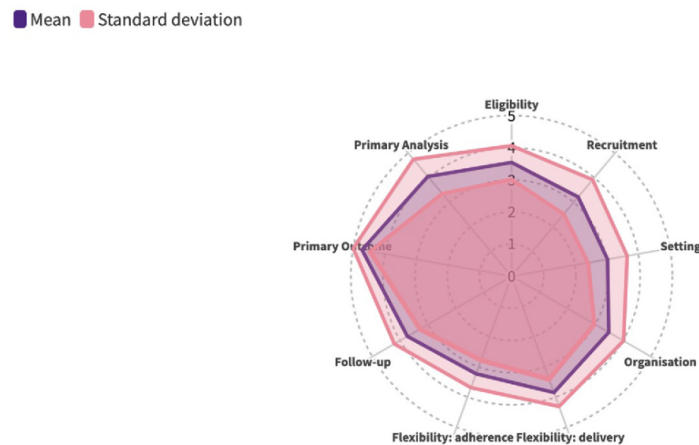


Fig. 2. Wheel diagram showing the mean PRECIS-2 score of the included studies.

4.

3.3. CONSORT assessment items

Forty-one per cent (16/39) of the trials declared being reported according to the CONSORT statement. Still, only one study [36] reported their findings adhering to the CONSORT extension for pragmatic trials. One out of three trials of the sample complied with more than 70 % of CONSORT items (13/39), between 50 % and 70 % in 49 % of the trials (19/39) and lower than 50 % of the CONSORT items in 13 % of the sample (5/39). Compliance of reporting varied across as seen in Table 3, and some items varied depending on whether they were published before or after the CONSORT statement publication (Supplementary file 5). The data analysis from the pre- and post-CONSORT periods indicates that the reporting of several items improved after the publication of the CONSORT tool (see supplementary file 5 for details). However, this improvement does not impact the final analysed reporting result for each item. The items reported by all the included trials were a structured abstract (1 b), clear eligibility criteria (4a) and statistical methods (12a). Regarding the CONSORT extension for pragmatic trials, all the items were reported in more than 72 % (28/39) of the cases except for item 5d (extra resources added to usual settings to implement the intervention). Although the description of the intervention for each study group (item 5a) was good in more than 70 % (28/39) of the trials, the authors did not sufficiently report details of the intervention, such as different components of the intervention, the standardisation or individualisation of the intervention and the additional sources to resemble clinical practice. Moreover, less than 20 % (7/39) of the trials reported whether and how the adherence of participants or care providers was assessed. Furthermore, items 'harms' and 'trial registration number and where the protocol can be assessed' were not reported in almost 50 % (16/39) of the sample. Thirty per cent (12/39) of the trials did not report strategies to limit bias when blinding was not possible, mechanisms to implement the random sequence allocation, the way in which the sample size was calculated or the setting where the data was collected. Twenty-five per cent (10/39) of the trials did not report the generalizability of the trial results or explain why blinding was not possible.

3.4. Risk of bias assessment

The overall risk of bias is summarised in Fig. 3. Performance and detection bias were the domains with a higher risk of bias (scored high in

90 %, 35/39 and 77 %; 30/39 of included trials, respectively). Reporting bias was unclear in almost half of the sample (46 %; 18/39). Supplementary file 4 contains a table showing the complete RoB assessment.

3.5. Relationship between the pragmatism, the quality of reporting and the RoB

The results recently published about the pragmatic attitude of the MT pRCTs [17] revealed that 61 % (24/39) of the studies were moderately pragmatic, 36 % (14/39) highly pragmatic, and 3 % (1/37) slightly pragmatic. No relation was found between the highly pragmatic attitude and good reporting except for CONSORT item 25 (Sources of funding and other support) ($p = 0.005$). Similarly, no relation was found between the RoB and the pragmatic attitude of the studies.

A comparison of the studies with a low vs high/unclear RoB according to the compliance with CONSORT items is provided in Table 4. The results showed that the percentage of compliance with CONSORT items was higher in the trials with low RoB in the following domains: Random sequence generation, allocation concealment, incomplete outcome data, and selective reporting. The comparison between RoB and each CONSORT item is shown in Supplementary File 5.

4. Discussion

This review described methodological and reporting features among self-labelled pRCTs in MT. It was found that the reporting is suboptimal in MT pRCTs. Performance and detection biases are present in MT pRCTs, mostly due to the intrinsic difficulties of MT interventions to manage blinding. Moreover, half of the sample showed unclear selective reporting. A relation was found between trials with low RoB and compliance with CONSORT RG.

About 40 % of the sample declared adherence to any reporting guideline. Only one trial [36] reported using the CONSORT extension for pragmatic trials [69]. This suggests that its usage may not be prevalent despite recommendations within the field [70]. The overall report was suboptimal. These results align with those of other authors assessing the reporting of RCTs in MT [1] and rehabilitation [6]. However, reporting items in the methods and results sections were slightly better reported in our sample, e.g. sample size description, randomisation and sequence allocation, and results description (flowchart).

Certain considerations about specific, pragmatic reporting features are worth mentioning. A significant cornerstone for understanding a

Table 3

Percentage of studies across each item of the CONSORT checklist for non-pharmacological interventions and the extension for pragmatic trials, in MT trials (N = 39).

ITEM	Studies with the item reported	
	N	%
1. TITLE AND ABSTRACT		
1a. Identification as a randomised trial in the title	33	85
1 b. Structured summary of trial design, methods, results, and conclusions	39	100
2. BACKGROUND		
2a. Scientific background and explanation of the rationale ^b Describe the health or health service problem that the intervention is intended to address and other interventions that may commonly be aimed at this problem	37	95
2 b. Specific objectives or hypotheses	38	98
METHODS		
3. TRIAL DESIGN		
3a. Description of trial design (e.g. parallel, factorial) including allocation ratio	27	69
3 b. Important changes to methods after trial commencement (such as eligibility criteria), with reasons	7	18
4. PARTICIPANTS		
4a. Eligibility criteria for participants	39	100
^a When applicable, eligibility criteria for centres and for care providers		
^b Eligibility criteria should be explicitly framed to show the degree to which they include typical participants and/or, where applicable, typical providers (e.g. nurses), institutions (e.g. hospitals), communities (or localities e.g. towns) and settings of care (e.g. different healthcare financing systems)		
4 b. Settings and locations where the data were collected	29	74
5. INTERVENTIONS		
5a. The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	28	72
^a Precise details of both the experimental treatment and comparator		
5 b. Description of the different components of the interventions and, when applicable, description of the procedure for tailoring the interventions to individual participants	21	54
5c. Details of whether and how the interventions were standardised	22	56
5d. Describe extra resources added to (or resources removed from) usual settings in order to implement the intervention. Indicate whether efforts were made to standardise the intervention or whether the intervention and its delivery were allowed to vary between participants, practitioners or study sites	15	38
5e. Details of whether and how adherence of care providers to the protocol was assessed or enhanced	7	17
5f. Details of whether and how adherence of participants to interventions was assessed or enhanced	6	15
6. OUTCOMES		
6a. Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	38	97
^b Explain why the chosen outcomes and, when relevant, the length of follow-up are considered important to those who will use the results of the trial		
6 b. Any changes to trial outcomes after the trial commenced, with reasons	3	8
7. SAMPLE SIZE		
7a. How sample size was determined	29	74
^b If calculated using the smallest difference considered important by the target decision-maker audience (the minimally important difference), then report where this difference was obtained		
7 b. When applicable, explanation of any interim analyses and stopping guidelines	2	5
8. SEQUENCE GENERATION		
8a. The method used to generate the random allocation sequence	35	90
8 b. Type of randomisation; details of any restriction (such as blocking and block size)	30	77
9. ALLOCATION CONCEALMENT		
9. The mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	28	72
10. IMPLEMENTATION		

Table 3 (continued)

ITEM	Studies with the item reported	
	N	%
10. Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	26	67
11. BLINDING		
11a. If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how	30	77
^b If blinding was not done, or was not possible, explain why		
11 b. If relevant, a description of the similarity of interventions	23	59
11c. If blinding was not possible, description of any alternative attempts to limit bias	27	69
12. STATISTICAL METHODS		
12a. Statistical methods used to compare groups for primary and secondary outcomes	39	100
^a When applicable, details of whether and how the clustering by care providers or centres was addressed		
12 b. Methods for additional analyses, such as subgroup analyses and adjusted analyses	21	54
RESULTS		
13. PARTICIPANTS FLOW (a diagram is strongly recommended)		
13a. For each group, the numbers of participants who were randomly assigned, received the intended treatment and were analysed for the primary outcome	34	87
^a The number of care providers or centres performing the intervention in each group and the number of patients treated		
13 b. For each group, losses and exclusions after randomisation, together with reasons	34	87
13c. For each group, the delay between randomisation and the initiation of the intervention	5	13
13d. Details of the experimental treatment and comparator as they were implemented	4	10
14. RECRUITMENT		
14a. Dates defining the periods of recruitment and follow-up	29	74
14 b. Why the trial ended or was stopped	37	95
15. BASELINE DATA		
15. A table showing baseline demographic and clinical characteristics for each group	37	95
^a When applicable, a description of care providers (case volume, qualification, expertise, etc.) and centres (volume) in each		
16. NUMBERS ANALYSED		
16. For each group, the number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	35	90
17. OUTCOMES AND ESTIMATION		
17a. Estimated effect size and its precision (such as 95 % confidence interval)	36	92
17 b. For binary outcomes, the presentation of both absolute and relative effect sizes is recommended	10	26
18. ANCILLARY ANALYSIS		
18. Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing prespecified from exploratory	15	38
19. HARMS		
19a. All significant harms or unintended effects in each group	22	56
20. TRIAL LIMITATIONS, ADDRESSING SOURCES OF POTENTIAL BIAS, IMPRECISION, AND, IF RELEVANT, MULTIPLICITY OF ANALYSES		
^a In addition, take into account the choice of the comparator, lack of or partial blinding, and unequal expertise of care providers or centres in each group		
21. Generalizability (external validity, applicability) of the trial findings	29	74
^a Generalizability (external validity) of the trial findings according to the intervention, comparators, patients, and care providers and centres involved in the trial		
^b Describe key aspects of the setting which determined the trial results. Discuss possible differences in other settings where clinical traditions, health service organisation, staffing, or resources may vary from those of the trial		
22. Interpretation consistent with results, balancing benefits and harms and considering other relevant evidence	32	82
23. Registration number and name of trial registry	25	64

(continued on next page)

Table 3 (continued)

ITEM	Studies with the item reported	
	N	%
24. Where the full trial protocol can be accessed, if available	23	59
25. Sources of funding and other support (such as supply of drugs), role of funders	32	82

^a Items from the CONSORT extension for non-pharmacological interventions.
^b Items from the CONSORT extension for pragmatic trials.

trial's pragmatism is how the intervention has been performed and its characteristics. This is particularly relevant in a non-pharmacological complex intervention such as MT [71]. In our sample, general compliance with item 5 of CONSORT (Interventions) was generally poor. Research in this area has previously identified the problem of inadequate reporting of interventions and lack of improvement over time [1]. Although general descriptions of the interventions were provided, details such as the procedure for tailoring the intervention to individual participants and how the adherence of participants and providers was assessed were present in less than half of the sample. These points are relevant in pRCTs and essential to understanding the study context for readers unfamiliar with the intervention and ensuring the intervention's reproducibility [72]. A protocolised intervention (because of the trial requirements) avoids biases and facilitates and allows trial replication [73]. However, it contraposes MT interventions that require a person-centred approach. The treatment can differ between patients because of the high variation among the same condition. For example,

non-specific lumbar pain can have multiple aetiology depending on each patient, thus requiring individual approaches for each person and different manual techniques according to the evolution of the condition [74]. Although MT interventions in pRCTs might be performed at the therapist's discretion, reporting the information about how the intervention will be tailored to individual participants will facilitate trial replication and directly increase the methodological quality of the trial [73,75]. In fact, this item (tailoring of the intervention) is included in the TIDIER guideline. However, no studies mentioned using the TIDIER checklist [76] or other reporting tools to describe the intervention. Despite several calls to improve the reporting of the intervention [76–78], a recent overview of reviews of healthcare research reports found that it was inadequate [79]. The comprehensive reporting of any research intervention in clinical trials is essential to assess the relevance of the findings in the context of routine practice [71,80].

Our sample showed good reporting of the way in which blinding was used and the strategies used to limit the bias when blinding was not possible. This is highly recommended in the CONSORT extension for non-pharmacological trials [28,81] and the CONSORT extension for pragmatic trials [69]. However, due to the difficulties in implementing effective blinding strategies in MT trials [73,75], our RoB assessment found performance and detection biases in almost the entire sample. This fact, which is very common in MT literature [73,75,82], greatly affects the perceived quality of any RCT and the credibility of the results. In the face of this situation, some authors could question the usefulness of the RoB tool to assess the quality of the trial evaluating a complex non-drug intervention such as MT, especially if the trial has been designed with a clear, pragmatic attitude and follows their premises. In fact, whether blinding should be performed in pRCTs is a subject of

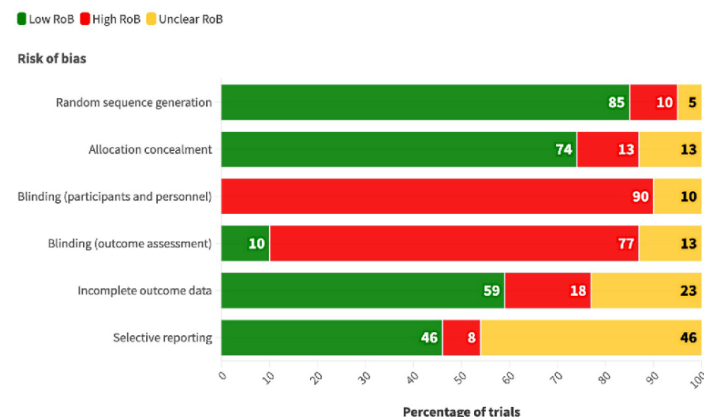


Fig. 3. Risk of Bias assessment (n = 39).

Table 4

Comparison between studies with low vs high/unclear RoB according to the compliance with CONSORT items.

Domains	Number of studies with a high/unclear risk of bias	% Compliance in CONSORT	Number of studies with a low risk of bias	% Compliance in CONSORT	P-value
Random sequence generation	7	46.7 (42.8–51.1)	32	68.9 (62.2–78.8)	<0.001
Allocation concealment	11	48.9 (46.2–58.9)	28	68.9 (64.4–78.9)	<0.001
Blinding (participants and personnel)	39	66.7 (56.7–73.3)	0	–	–
Blinding (outcome assessment)	33	66.7 (57.8–77.8)	6	67.8 (58.4–68.9)	0.598
Incomplete outcome data	17	64.4 (46.7–68.9)	22	68.9 (60.5–81.1)	0.049
Selective reporting	22	58.9 (47.3–66.7)	17	73.3 (68.9–82.2)	<0.001

Note: Data are presented as median and interquartile range. The number of studies in the low risk of bias or high/unclear risk of bias category varies in each domain of the RoB tool.

discussion [83–85]; instead, open-label trials and masking the external assessor alternatives have been proposed to resemble clinical practice better [82]. Moreover, patient-reported outcomes constitute a key feature of pRCTs. However, if patients have not been blinded, masking the outcome assessors becomes impossible, leading again to some sort of bias. The IMMPACT statement recently published about research objectives and considerations on pragmatic trials of pain intervention proposed different strategies to limit bias in pRCTs, such as nonblinded comparative effectiveness trials, blinding to study hypothesis, Zelen designs or use of data capture systems for patient-reported outcomes [20].

Additionally, there is an urgent call for MT pRCTs to register research protocols and to report whether the protocol exists and where it is to be found. Reporting bias was unclear in half of our sample for this reason.

Looking at the reporting and RoB assessment results, we concur that good reporting is necessary to assess the risk of bias in trials properly. This is essential to determine whether any bias influences the interventions' treatment effect [7]. In fact, in MT and physical rehabilitation, recent meta-research studies have shown a high correlation between good reporting and the quality of the studies [1,6].

5. Strengths and limitations

While our sample may not be comprehensive in covering all MT modalities, it does include all pRCTs published in the MT field from inception to January 2024.

We recognise that the reviewer group was sizeable, which could have contributed to heterogeneity in the data extraction process and influenced the outcomes. To reduce this limitation, we conducted a pilot of the data extraction form and had extensive discussions on the ratings prior to commencing the study. Additionally, all reviewers were paired with the IP to discuss the scores and minimise subjectivity. To describe the pragmatism of the included studies, we have used the methodology described by other authors in their studies [16,86–92]. While this methodology lacks scientific validation, the authors deemed it worthwhile to replicate it to facilitate a comparison between manual therapy and other healthcare disciplines. Using available tools to assess the risk of bias in designs with the particularities of a pragmatic trial remains a challenge. However, it is important to emphasize that the most widely accepted approaches for conducting this assessment include domains that have empirically been shown to impact the internal validity of studies. Moreover, we conducted the review in accordance with the main methodological guidelines of Cochrane. The interest to assess the health research methodology is evolving. The homogenization and consensus in the methodological proposals for this type of study [93,94] should enable their implementation in a systematic and rigorous manner.

6. Conclusions

The standards of the reporting are suboptimal in MT pRCTs. Also, such studies are commonly affected by performance and detection bias due to the difficulties related to blinding strategies. Overall, neither the quality of the reporting nor the RoB is related to a higher or lower pragmatic attitude. Trials with low RoB have shown higher compliance with CONSORT reporting guidelines.

Ethical approval

Not applicable.

Funding

Not applicable.

CRediT authorship contribution statement

S. Roura: Writing – review & editing, Writing – original draft, Methodology, Investigation, Data curation, Conceptualization. **G. Alvarez:** Writing – review & editing, Supervision, Methodology, Investigation, Conceptualization. **D. Hohenschurz-Schmidt:** Writing – review & editing, Investigation. **I. Solà:** Writing – review & editing, Supervision, Methodology. **R. Núñez-Cortés:** Writing – review & editing, Methodology, Investigation. **J. Bracchiglione:** Writing – review & editing, Investigation. **C. Fernández-Jané:** Investigation. **J. Phalip:** Investigation. **I. Gich:** Formal analysis. **M. Sitjà-Rabert:** Writing – review & editing, Supervision. **G. Urrútia:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Sònia Roura is a PhD student in Biomedical Research Methodology and Public Health in the Medical Department of the Universitat Autònoma de Barcelona, Barcelona, Spain.

We would like to thank Roberto Acosta for his support in the first stages of the project.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ijosm.2025.100754>.

ABBREVIATIONS

MT	Manual Therapy
pRCT	Pragmatic Randomised Controlled Trial
RoB	Risk of Bias
CI	Complex Intervention
RG	Reporting Guideline

References

- [1] Alvarez G, Solà I, Sitjà-Rabert M, Fort-Vanmeerhaeghe A, Gich I, Fernández C, et al. A methodological review revealed that reporting of trials in manual therapy has not improved over time. *J Clin Epidemiol* 2020;121:32–44. <https://doi.org/10.1016/j.jclinepi.2020.01.006>.
- [2] Riley SP, Swanson BT, Sawyer SE, Brismée J-M. Is research quality in orthopedic manual therapy trials stagnating? Reflections and pathways for improving research quality and advance our profession. *J Man Manip Ther* 2016;24:239–40. <https://doi.org/10.1080/10669817.2016.1253561>.
- [3] González GZ, Moseley AM, Maher CG, Nascimento DP, Costa L, da CM, Costa LO. Methodologic quality and statistical reporting of physical therapy randomized controlled trials relevant to musculoskeletal conditions. *Arch Phys Med Rehabil* 2018;99:129–36. <https://doi.org/10.1016/j.apmr.2017.08.485>.
- [4] Núñez-Cortés R, Alvarez G, Pérez-Bracchiglione J, Cabanas-Valdés R, Calvo-Sanz J, Bonfill X, et al. Reporting results in manual therapy clinical trials: a need for improvement. *Int J Osteopath Med* 2021. <https://doi.org/10.1016/j.ijosm.2021.06.002>.
- [5] Alvarez G, Núñez-Cortés R, Solà I, Sitjà-Rabert M, Fort-Vanmeerhaeghe A, Fernández C, et al. Sample size, study length, and inadequate controls were the most common self-acknowledged limitations in manual therapy trials: a methodological review. *J Clin Epidemiol* 2021;130:96–106. <https://doi.org/10.1016/j.jclinepi.2020.10.018>.
- [6] Innocenti T, Giagio S, Salvioli S, Feller D, Minnucchi S, Brindisino F, et al. Completeness of reporting is suboptimal in randomized controlled trials published in rehabilitation journals, with trials with low risk of bias displaying better reporting: a meta-research study. *Arch Phys Med Rehabil* 2022. <https://doi.org/10.1016/j.apmr.2022.01.156>.
- [7] Arienti C, Armijo-Olivo S, Ferrero G, Feys P, Hoogbeem T, Kiekens C, et al. The influence of bias in randomized controlled trials on rehabilitation intervention effect estimates: what we have learned from meta-epidemiological studies. *Eur J*

- Phys Rehabil Med 2024;60:135–44. <https://doi.org/10.23736/S1973-9087.23.08310-7>.
- [8] Godwin M, Ruhl and I, Casson I, MacDonald S, Delva D, Birtwhistle R, et al. Pragmatic controlled clinical trials in primary care: the struggle between external and internal validity. *BMC Med Res Methodol* 2003;3:28. <https://doi.org/10.1186/1471-2288-3-28>.
 - [9] Ford I, Norrie J. Pragmatic trials. *N Engl J Med* 2016;375:454–63. <https://doi.org/10.1056/nejr1510059>.
 - [10] Dieppe P. Complex interventions. *Musculoskeletal Care* 2004;2:180–6. <https://doi.org/10.1002/msc.69>.
 - [11] Walach H, Falkenberg T, Esmenbø V, Lewith G, Jonas WB. Circular instead of hierarchical: methodological principles for the evaluation of complex interventions. *BMC Med Res Methodol* 2006;6:29. <https://doi.org/10.1186/1471-2288-6-29>.
 - [12] Minary I, Trompette J, Kivits J, Cambon L, Tarquinio G, Alla F. Which design to evaluate complex interventions? Toward a methodological framework through a systematic review. *BMC Med Res Methodol* 2019;19:92. <https://doi.org/10.1186/s12874-019-0736-6>.
 - [13] Kerry R. Expanding our perspectives on research in musculoskeletal science and practice. *Musculoskeletal Sci Pract* 2017;32:114–9. <https://doi.org/10.1016/j.msksp.2017.10.004>.
 - [14] Daniel Maddox C, Subialka JA, Young JL, Rhon DI. TITLE: over half of clinical trials of mobilization and manipulation for patients with low back pain may have limited real-world applicability. A Systematic Review of 132 Clinical Trials n.d..
 - [15] Nicholls SG, Zwarenstein M, Hey SP, Girardeau B, Campbell MK, Taljaard M. The importance of decision intent within descriptions of pragmatic trials. *J Clin Epidemiol* 2020;125:30–7. <https://doi.org/10.1016/j.jclinepi.2020.04.030>.
 - [16] Del-Ré R, Janiaud P, Ioannidis JPA. Real-world evidence: how pragmatic are randomized controlled trials labeled as pragmatic? *BMC Med* 2018;16:49. <https://doi.org/10.1186/s12916-018-1038-2>.
 - [17] Roura S, Alvarez G, Hohenschurz-Schmidt D, Solà I, Núñez-Cortés R, Bracciglion J, et al. Lack of pragmatic attitude of self-labeled pragmatic trials on manual therapy: a methodological review. *BMC Med Res Methodol* 2024;24. <https://doi.org/10.1186/s12874-024-02393-1>.
 - [18] Riley SP, Swanson B, Brismée J-M, Sawyer SF. A systematic review of orthopaedic manual therapy randomized clinical trials quality. *J Man Manip Ther* 2016;24:241–52. <https://doi.org/10.1080/10669817.2015.1119372>.
 - [19] Mosely AM, Ekins MR, Janer-Duncan L, Hush JM. The quality of reports of randomized controlled trials varies between disciplines of physiotherapy. *Physiother Can* 2014;66:36–43. <https://doi.org/10.3138/ptc.2012-68>.
 - [20] Hohenschurz-Schmidt DJ, Cherkin D, Rice ASG, Dworkin RH, Turk DC, McDermott MP, et al. Research objectives and general considerations for pragmatic clinical trials of pain treatments: IMMPACT statement. *Pain* 2023. <https://doi.org/10.1097/j.pain.0000000000002888>.
 - [21] Zwarenstein M, Treweek S, London K. PRECIS-2 helps researchers design more applicable RCTs while CONSORT Extension for Pragmatic Trials helps knowledge users decide whether to apply them. *J Clin Epidemiol* 2017;84:27–9. <https://doi.org/10.1016/j.jclinepi.2016.10.010>.
 - [22] Innocenti T, Salvioli S, Giagio S, Fellet D, Cartabellotta M, Chiarotto A. Declaration of use and appropriate use of reporting guidelines in high-impact rehabilitation journals is limited: a meta-research study. *J Clin Epidemiol* 2021;131:43–50. <https://doi.org/10.1016/j.jclinepi.2020.11.010>.
 - [23] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Rev Esp Cardiol* 2021;74:790–9. <https://doi.org/10.1016/j.rec.2021.07.010>.
 - [24] Sacristán JA, Dilla T. Pragmatic trials revisited: applicability is about individualization. *J Clin Epidemiol* 2018;99:164–6. <https://doi.org/10.1016/j.jclinepi.2018.02.003>.
 - [25] Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. *Syst Rev* 2016;5:210. <https://doi.org/10.1186/s13643-016-0384-4>.
 - [26] Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *Br Med J* 2010;340:e869. <https://doi.org/10.1136/bmj.e869>.
 - [27] Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, et al. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *Br Med J* 2008;337:a2390. <https://doi.org/10.1136/bmj.a2390>.
 - [28] Boutron I, Altman DG, Moher D, Schulz KF, Ravaut P, et al., CONSORT NPT Group. CONSORT statement for randomized trials of nonpharmacologic treatments: a 2017 update and a CONSORT extension for nonpharmacologic trial abstracts. *Ann Intern Med* 2017;167:40–7. <https://doi.org/10.7326/M17-0046>.
 - [29] Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *Br Med J* 2011;343:d5928. <https://doi.org/10.1136/bmj.d5928>.
 - [30] Lee J, Cho J-H, Kim K-W, Lee J-H, Kim M-R, Kim J, et al. Chuna manual therapy vs usual care for patients with nonspecific chronic neck pain: a randomized clinical trial. *JAMA Netw Open* 2021;4:e2113757. <https://doi.org/10.1001/jamanetworkopen.2021.13757>.
 - [31] Nguyen AP, Pitance L, Mahaudens P, Detrembleur C, David Y, Hall T, et al. Effects of mulligan mobilization with movement in subacute lateral ankle sprains: a pragmatic randomized trial. *J Man Manip Ther* 2021;29:341–52. <https://doi.org/10.1080/10669817.2021.1889165>.
 - [32] Groisman S, Malysz T, de Souza da Silva L, Rocha Ribeiro Sanches T, Camargo Bragante K, Locatelli F, et al. Osteopathic manipulative treatment combined with exercise improves pain and disability in individuals with non-specific chronic neck pain: a pragmatic randomized controlled trial. *J Bodyw Mov Ther* 2019;24:189–95. <https://doi.org/10.1016/j.jbmt.2019.11.002>.
 - [33] Wilkey A, Gregory M, Byfield D, McCarthy PW. A comparison between chiropractic management and pain clinic management for chronic low-back pain in a national health service outpatient clinic. *J Alternative Compl Med* 2008;14:465–73. <https://doi.org/10.1089/acm.2007.0796>.
 - [34] Finch P, Bessonnette S. A pragmatic investigation into the effects of massage therapy on the self-efficacy of multiple sclerosis clients. *J Bodyw Mov Ther* 2014;18:11–6. <https://doi.org/10.1016/j.jbmt.2013.04.001>.
 - [35] Griswold D, Learman K, O'Halloran B, Côté J. A preliminary study comparing the use of cervical/upper thoracic mobilization and manipulation for individuals with mechanical neck pain. *J Man Manip Ther* 2015;23:75–83. <https://doi.org/10.1179/2042618614Y.0000000095>.
 - [36] Cross J, Hender F, Barton G, Clark A, Shepherson L, Blyth A, et al. A randomised controlled equivalence trial to determine the effectiveness and cost-utility of manual chest physiotherapy techniques in the management of exacerbations of chronic obstructive pulmonary disease (MATREX). *Health Technol Assess* 2010;14. <https://doi.org/10.3310/hta14230>.
 - [37] Poole H, Glenn S, Murphy P. A randomised controlled study of reflexology for the management of chronic low back pain. *Eur J Pain* 2007;11:878–87. <https://doi.org/10.1016/j.ejpain.2007.01.006>.
 - [38] Sharp DM, Walker MB, Chaturvedi A, Upadhyay S, Hamid A, Walker AA, et al. A randomised, controlled trial of the psychological effects of reflexology in early breast cancer. *Eur J Cancer* 2010;46:312–22. <https://doi.org/10.1016/j.ejca.2009.10.006>.
 - [39] Stockkendahl MJ, Christensen HW, Vach W, Høiland-Andersen PE, Haghihi T, Hartvigsen J. A randomized clinical trial of chiropractic treatment and self-management in patients with acute musculoskeletal chest pain: 1-year follow-up. *J Man Manip Ther* 2012;35:254–62. <https://doi.org/10.1016/j.jmpt.2012.04.003>.
 - [40] Goertz CM, Long CR, Hondras MA, Petri R, Delgado R, Lawrence DJ, et al. Adding chiropractic manipulative therapy to standard medical care for patients with acute low back pain: results of a pragmatic randomized comparative effectiveness study. *Spine* 2013;38:627–34. <https://doi.org/10.1097/BRS.0b013e3182773357>.
 - [41] Attias S, Sivan K, Avneri O, Sagee A, Ben-Arye E, Grünberg O, et al. Analgesic effects of reflexology in patients undergoing surgical procedures: a randomized controlled trial. *J Alternative Compl Med* 2018;24:809–15. <https://doi.org/10.1089/acm.2017.0167>.
 - [42] Park S-Y, Hwang E-H, Cho J-H, Kim K-W, Ha I-H, Kim M-R, et al. Comparative effectiveness of chuna manipulative therapy for non-acute lower back pain: a multi-center, pragmatic, randomized controlled trial. *J Clin Med Res* 2020;9. <https://doi.org/10.3390/jcm9010144>.
 - [43] Hay EM, Mullis R, Lewis M, Vohora K, Main GJ, Watson P, et al. Comparison of physical treatments versus a brief pain-management programme for back pain in primary care: a randomised clinical trial in physiotherapy practice. *Lancet* 2005;365:2024–30. [https://doi.org/10.1016/S0140-6736\(05\)66696-2](https://doi.org/10.1016/S0140-6736(05)66696-2).
 - [44] Dissing KB, Hartvigsen J, Wedderkopp N, Hestbak L. Conservative care with or without manipulative therapy in the management of back and/or neck pain in Danish children aged 9–15: a randomised controlled trial nested in a school-based cohort. *BMJ Open* 2018;8:e021358. <https://doi.org/10.1136/bmjopen-2017-021358>.
 - [45] Wyatt K, Edwards V, Franck L, Britten N, Creanor S, Maddick A, et al. Cranial osteopathy for children with cerebral palsy: a randomised controlled trial. *Arch Dis Child* 2011;96:505–12. <https://doi.org/10.1136/adc.2010.199877>.
 - [46] Goertz CM, Long CR, Vining RD, Pohlman KA, Walter J, Coulter I. Effect of usual medical care plus chiropractic care vs usual medical care alone on pain and disability among us service members with low back pain: a comparative effectiveness clinical trial. *JAMA Netw Open* 2018;1:e180105. <https://doi.org/10.1001/jamanetworkopen.2018.0105>.
 - [47] Castien RE, van der Windt DAWM, Grooten A, Dekker J. Effectiveness of manual therapy for chronic tension-type headache: a pragmatic, randomised, clinical trial. *Cephalalgia* 2011;31:133–43. <https://doi.org/10.1177/0333102410377362>.
 - [48] Dziedzic K, Hill J, Lewis M, Sim J, Daniels J, Hay EM. Effectiveness of manual therapy or pulsed shortwave diathermy in addition to advice and exercise for neck disorders: a pragmatic randomized controlled trial in physical therapy clinics. *Arthritis Rheum* 2005;53:214–22. <https://doi.org/10.1002/art.21087>.
 - [49] Mafetoni RR, Shimo AKK. Effects of acupressure on progress of labor and cesarean section rate: randomized clinical trial. *Rev Saude Publica* 2015;49:9. <https://doi.org/10.1590/s0034-8910.2015049005407>.
 - [50] Miller JE, Newell D, Bolton JE. Efficacy of chiropractic manual therapy on infant colic: a pragmatic single-blind, randomized controlled trial. *J Man Manip Ther* 2012;35:600–7. <https://doi.org/10.1016/j.jmpt.2012.09.010>.
 - [51] Walach H, Güthlin C, König M. Efficacy of massage therapy in chronic pain: a pragmatic randomized trial. *J Alternative Compl Med* 2003;9:837–46. <https://doi.org/10.1089/107555303771952181>.
 - [52] Harper B, Steinbeck L, Aron A. Fascial manipulation vs. standard physical therapy practice for low back pain diagnoses: a pragmatic study. *J Bodyw Mov Ther* 2019;23:115–21. <https://doi.org/10.1016/j.jbmt.2018.10.007>.
 - [53] Bergman GJ, Winters JC, Groenier KH, Meyboom-de Jong B, Postema K, van der Heijden GJ. Manipulative therapy in addition to usual care for patients with shoulder complaints: results of physical examination outcomes in a randomized controlled trial. *J Man Manip Ther* 2010;33:96–101. <https://doi.org/10.1016/j.jmpt.2009.12.004>.
 - [54] Groeneweg R, van Assen L, Kropman H, Leopold H, Mulder J, Smits-Engelsman BCM, et al. Manual therapy compared with physical therapy in patients

- with non-specific neck pain: a randomized controlled trial. *Chiropr Man Ther* 2017; 25:12. <https://doi.org/10.1186/s12998-017-0141-3>.
- [55] Hoving. Manual Therapy. Physical therapy, or continued care by the general practitioner for patients with neck pain long-term results from a pragmatic randomized clinical trial. 2006.
- [56] Lilje S, Friberg H, Wykman A, Skilgate E. Naprapathic manual therapy or conventional orthopedic care for outpatients on orthopedic waiting lists: a pragmatic randomized controlled trial. *Clin J Pain* 2010;26:602–10. <https://doi.org/10.1097/AJP.0b013e3181d71ebd>.
- [57] Schweda F, Rother K, Rother D, Ruetz M, Resch K-L. Osteopathic manipulative therapy in women with postpartum low back pain and disability: a pragmatic randomized controlled trial. *J Am Osteopath Assoc* 2015;115:416–25. <https://doi.org/10.7556/jaoa.2015.087>.
- [58] Williams NH, Wilkinson C, Russell I, Edwards RT, Hibbs R, Linek P, et al. Randomized osteopathic manipulation study (ROMANS): pragmatic trial for spinal pain in primary care. *Fam Pract* 2003;20:662–9. <https://doi.org/10.1093/fampra/cmg007>.
- [59] Gemmell H, Miller P. Relative effectiveness and adverse effects of cervical manipulation, mobilisation and the activator instrument in patients with sub-acute non-specific neck pain: results from a stopped randomised trial. *Chiropr Osteopathy* 2010;18. <https://doi.org/10.1186/1746-1340-18-20>.
- [60] Castro-Sánchez AM, Lara-Palomo IG, Matarán-Peñarocha GA, Fernández-de-las-Peñas C, Saavedra-Hernández M, Cleland J, et al. Short-term effectiveness of spinal manipulative therapy versus functional technique in patients with chronic nonspecific low back pain: a pragmatic randomized controlled trial. *Spine J* 2016; 16:302–12. <https://doi.org/10.1016/j.spinee.2015.08.057>.
- [61] Bronfort G, Hondras MA, Schulz CA, Evans RL, Long CR, Grimmer R. Spinal manipulation and home exercise with advice for subacute and chronic back-related leg pain: a trial with adaptive allocation. *Ann Intern Med* 2014;161:381–91. <https://doi.org/10.7326/M14-0006>.
- [62] Evans R, Haas M, Schulz C, Leininger B, Hanson L, Bronfort G. Spinal manipulation and exercise for low back pain in adolescents: a randomized trial. *Pain* 2018;159: 1297–307. <https://doi.org/10.1097/j.pain.0000000000001211>.
- [63] Georgoudis G, Petah B, Nikolaidis P, Damigos D. The effect of myofascial release and microwave diathermy combined with acupuncture versus acupuncture therapy in tension-type headache patients: a pragmatic randomized controlled trial. *Physiother Res Int* 2018;23:e1700. <https://doi.org/10.1002/pt.1700>.
- [64] Skilgate E, Bohman T, Holm LW, Vingård E, Alftedsson L. The long-term effects of naprapathic manual therapy on back and neck pain - results from a pragmatic randomized controlled trial. *BMC Musculoskelet Disord* 2010;11:26. <https://doi.org/10.1186/1471-2474-11-26>.
- [65] Eldred A, Jensen I, Lohela-Karlsson M, Hagberg J, Leboeuf-Yde C, Kongsted A, et al. The Nordic Maintenance Care program: effectiveness of chiropractic maintenance care versus symptom-guided treatment for recurrent and persistent low back pain—a pragmatic randomized controlled trial. *PLoS One* 2018;13: e0203029. <https://doi.org/10.1371/journal.pone.0203029>.
- [66] Lim K-T, Hwang E-H, Cho J-H, Jung J-Y, Kim K-W, Ha I-H, et al. Comparative effectiveness of Chuna manual therapy versus conventional usual care for non-acute low back pain: a pilot randomized controlled trial. *Trials* 2019;20:216. <https://doi.org/10.1186/s13063-019-3302-y>.
- [67] Kim B-J, Park A-L, Hwang M-S, Heo I, Park S-Y, Cho J-H, et al. Comparative effectiveness and safety of concomitant treatment with chuna manual therapy and usual care for whiplash injuries: a multicenter randomized controlled trial. *Int J Environ Res Public Health* 2022;19. <https://doi.org/10.3390/ijerph191710678>.
- [68] Lynen A, Schönlitz M, Vahle M, Jäkel A, Rütz M, Schwerla F. Osteopathic treatment in addition to standard care in patients with Gastroesophageal Reflux Disease (GERD) - a pragmatic randomized controlled trial. *J Bodyw Mov Ther* 2022;29: 223–31. <https://doi.org/10.1016/j.jbmt.2021.09.017>.
- [69] Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, et al. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *Br Med J* 2008;337:a2390.
- [70] Eldred WG, Munk N. Using the pragmatic-explanatory continuum indicator summary (PRECIS) model in clinical research: application to refine a practice-based research network (PBRN) study. *J Am Board Fam Med* 2014;27:846–54. <https://doi.org/10.3122/jabfm.2014.06.140042>.
- [71] Alvarez G, Cerritelli F, Urrutia G. Using the template for intervention description and replication (TIDieR) as a tool for improving the design and reporting of manual therapy interventions. *Man Ther* 2016;24:85–9. <https://doi.org/10.1016/j.math.2016.03.004>.
- [72] Leech JB, Owen WE, Young JL, Rhon DI. Incomplete reporting of manual therapy interventions and a lack of clinician and setting diversity in clinical trials for neck pain limits replication and real-world translation. A scoping review. *J Man Manip Ther* 2022;1–10. <https://doi.org/10.1080/10669817.2022.2113295>.
- [73] Koes BW. How to evaluate manual therapy: value and pitfalls of randomized clinical trials. *Man Ther* 2004;9:183–4. <https://doi.org/10.1016/j.math.2004.04.002>.
- [74] Bishop MD, Torres-Cueco R, Gay CW, Lynch-Gibbs E, Beneciuk JM, Bialosky JE. What effect can manual therapy have on a patient's pain experience? *Pain Manag* 2015;5:455–64. <https://doi.org/10.2217/pmt.15.39>.
- [75] Wade DT, Smeets RJEM, Verbunt JA. Research in rehabilitation medicine: methodological challenges. *J Clin Epidemiol* 2010;63:699–704. <https://doi.org/10.1016/j.jclinepi.2009.07.010>.
- [76] Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *Br Med J* 2014;348:g1687. <https://doi.org/10.1136/bmj.g1687>.
- [77] Hoffmann TC, Eruti C, Glasziou PP. Poor description of non-pharmacological interventions: analysis of consecutive sample of randomised trials. *Br Med J* 2013; 347. <https://doi.org/10.1136/bmj.f3755>.
- [78] Glasziou P, Meats E, Heneghan C, Shepperd S. What is missing from descriptions of treatment in trials and reviews? *Br Med J* 2008;336:1472–4. <https://doi.org/10.1136/bmj.39590.732037.47>.
- [79] Dijkers MP. Overview of reviews using the template for intervention description and replication (TIDieR) as a measure of trial intervention reporting quality. *Arch Phys Med Rehabil* 2021;102:1623–32. <https://doi.org/10.1016/j.apmr.2020.09.397>.
- [80] Yamato T, Maher C, Saragiotto B, Moseley A, Hoffmann T, Elkins M, et al. Improving completeness and transparency of reporting in clinical trials using the template for intervention description and replication (TIDieR) checklist will benefit the physiotherapy profession. *J Man Manip Ther* 2016;24:183–4. <https://doi.org/10.1080/10669817.2016.1210343>.
- [81] Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P. Methods and processes of the CONSORT group: example of an extension for trials assessing nonpharmacologic treatments. *Ann Intern Med* 2008. <https://doi.org/10.7326/0003-4819-148-4-200802190-00008-wj1>.
- [82] Villamar MF, Contreras VS, Kuntz RE, Fregni F. The reporting of blinding in physical medicine and rehabilitation randomized controlled trials: a systematic review. *J Rehabil Med* 2013;45:6–13. <https://doi.org/10.2340/16501977.1071>.
- [83] Dal-Ré R. The PRECIS-2 tool seems not to be useful to discriminate the degree of pragmatism of medicine masked trials from that of open-label trials. *Eur J Clin Pharmacol* 2021;77:539–46. <https://doi.org/10.1007/s00228-020-03030-8>.
- [84] Dal-Ré R. Pragmatic trials, blinding, placebo, and the usefulness of the PRECIS-2 tool. *Eur J Clin Pharmacol* 2021;77:1071–2. <https://doi.org/10.1007/s00228-020-03079-5>.
- [85] Zwarenstein M, Howie A. Blinding, pragmatism, and the PRECIS-2 tool for designing and assessing randomized trials. *Eur J Clin Pharmacol* 2021;77:1069–70. <https://doi.org/10.1007/s00228-020-03078-6>.
- [86] Janiaud P, Dal-Ré R, Ioannidis JPA. Assessment of pragmatism in recently published randomized clinical trials. *JAMA Intern Med* 2018;178:1278. <https://doi.org/10.1001/jamainternmed.2018.3321>.
- [87] Hohenschurz-Schmidt D, Kleykamp BA, Draper-Rodi J, Vollett J, Chan J, Ferguson M, et al. Pragmatic trials of pain therapies: a systematic review of methods. *Pain* 2021.
- [88] Sepehrvand M, Alemayehu W, Das D, Gupta AK, Gouda P, Ghimire A, et al. Trends in the explanatory or pragmatic nature of cardiovascular clinical trials over 2 decades. *JAMA Cardiol* 2019;4:1122–8. <https://doi.org/10.1001/jamacardio.2019.3604>.
- [89] Devos F, Foissac F, Bouazza N, Aned P-Y, Tréuyer J-M, Chappuy H. Study characteristics impacted the pragmatism of randomized controlled trial published in nursing: a meta-epidemiological study. *J Clin Epidemiol* 2019;116:18–25. <https://doi.org/10.1016/j.jclinepi.2019.07.017>.
- [90] Palakshappa JA, Gibbs KW, Lannan MT, Cranford AR, Taylor SP. Systematic review of the “pragmatism” of pragmatic critical care trials. *Critical Care Explorations* 2022;4:e0738. <https://doi.org/10.1097/CCE.0000000000000738>.
- [91] Taljaard M, Nicholls SG, Howie AH, Nix HP, Carroll K, Moon PM, et al. An analysis of published trials found that current use of pragmatic trial labels is uninformative. *J Clin Epidemiol* 2022. <https://doi.org/10.1016/j.jclinepi.2022.08.007>.
- [92] Rosas LG, Lv N, Azar K, Xiao L, Yank V, Ma J. Applying the pragmatic-explanatory continuum indicator summary model in a primary care-based lifestyle intervention trial. *Am J Prev Med* 2015;49:S208–14. <https://doi.org/10.1016/j.amepre.2015.05.011>.
- [93] Mbuagbaw L, Lawson DO, Puljak L, Allison DB, Thabane L. A tutorial on methodological studies: the what, when, how and why. *BMC Med Res Methodol* 2020;20:226. <https://doi.org/10.1186/s12874-020-01107-7>.
- [94] Gentles SJ, Charles C, Nicholas DB, Ploeg J, McKibbin KA. Reviewing the research methods literature: principles and strategies illustrated by a systematic overview of sampling in qualitative research. *Syst Rev* 2016;5:172. <https://doi.org/10.1186/s13643-016-0343-0>.

4.3 Resultados del tercer trabajo

4.3.1 Resultados de la búsqueda

La búsqueda identificó 3553 artículos. Tras cribar el título, el resumen y la disponibilidad del texto completo, la muestra final comprendió 306 estudios. Entre éstos, 39 se identificaron como pragmáticos autodenominados en el título y/o el resumen (EP), mientras que 267 fueron Ensayos Controlados Aleatorios de «efectividad» no autodenominados como pragmáticos (EE). No se identificó ningún artículo en TM con el término naturalista. De los 267 ensayos de efectividad, 35 eran ensayos controlados con placebo (EEplacebo), mientras que los 232 restantes empleaban otras intervenciones aceptadas o establecidas (EEotraitervención). Para equilibrar y comparar los tres grupos, los 232 EEotraitervención se aleatorizaron utilizando el generador aleatorio de Microsoft Excel para obtener un grupo de 39 ensayos (el mismo número de ensayos que en el grupo EPs). Tal y como se ha explicado en el apartado de métodos 3.3.3.

4.3.2 Comparación de las características principales de los estudios incluidos

La media de participantes fue significativamente menor en los EE que en los EPs. En los EPs, la intervención estudiada con mayor frecuencia fue una combinación de técnicas no protocolizada (54%; 21/39), mientras que en EEotraitervención fue una combinación estandarizada de técnicas (51%; 20/39) y una intervención con una técnica aislada no estandarizada en el 49% (17/35) de los EEplacebo. El diseño y la intervención control utilizados con mayor frecuencia en los EPs y EEotraitervención fueron un diseño de dos grupos comparados con otra intervención o “usual care” como grupo control. En los EE placebo, se utilizó más un diseño de dos grupos con un grupo de control placebo. Si bien no se realizó el enmascaramiento de los terapeutas en ninguno de los tres grupos, los EE obtuvieron tasas más altas de enmascaramiento de los evaluadores externos que

los EP. El entorno de los participantes difirió sustancialmente entre los tipos de estudios, siendo los estudios multicéntricos más frecuentes en los EPs que en los EE. La tabla 1 (de la publicación respectiva) describe la comparación de las principales características de la muestra.

Todos los estudios se distribuyeron de manera similar en cuanto a si se publicaron en revistas especializadas o no especializadas en el campo de la TM (46%;52 vs 54%;61). La guía CONSORT se reportó haber sido utilizada en el 41% (16/39) de los EP, el 31% (12/39) de los EEotraitervención y el 26% del EEplacebo (9/35).

Aunque el término “Pragmático” no apareció en el título ni en el resumen en los EE no autodenominados pragmáticos, el 64% (25/39) del EEotraitervención declaró tener una actitud pragmática respecto a la aplicabilidad de los resultados del estudio. Sólo el 31% (11/35) de los EE placebo lo hicieron. Las limitaciones reportadas principalmente por los autores en los diferentes tipos de estudios fueron el enmascaramiento de los participantes (17/14/12), la dificultad en generalizar los resultados (13/11/11) y el tamaño de la muestra (12/11/14). El control inadecuado (13/8/4) fue frecuente en los EPs y los Eeotraitervención, pero no en los EEplacebo. Las limitaciones en la duración de los estudios fueron frecuentes en los EE en general.

4.3.3 Evaluación retrospectiva del pragmatismo a través de la herramienta PRECIS-2

Al diferenciar entre ligeramente, moderado y altamente pragmático, el 60% de los estudios tuvieron una actitud pragmática moderada. La figura 3 (de la publicación respectiva) muestra la rueda PRECIS-2 según el tipo de estudio.

Con respecto a los EP, casi todos los dominios (excepto el Dominio 6) fueron calificados como pragmáticos en > 50% de los estudios. Para los EE, los dominios 1 (elegibilidad), 2 (reclutamiento) y 8 (variable principal) se calificaron como pragmáticos en más del 50%

de los estudios. Los dominios 8 (variable principal) y 9 (análisis primario) fueron calificados como pragmáticos en más del 50% de todos los estudios. El dominio 6 (adherencia a la intervención) fue el dominio peor informado en todos los tipos de estudios. El dominio 3 (entorno) se calificó como explicativo en un alto porcentaje de los EE (82%). La flexibilidad en la intervención (Dominio 5) fue pragmática en más del 50% de los EPs y explicativa en más del 50% de los EE.

4.3.4 Evaluación del reporte a través del cumplimiento de los ítems de la guía CONSORT

Casi un tercio de los estudios (37/113) se reportaron según la guía CONSORT, el 41 % (16/39) de los EP, el 31 % (12/39) de los EEotraitervención y el 26 % (12/35) de los EE placebo.

Uno de cada cuatro estudios de la muestra reportó detalles relacionados con más del 70% de los ítems CONSORT (13/39 del EP, 9/39 del EEotraitervención y 13/32 del EE placebo), entre el 50% y el 70% en 41 % de los estudios (46/113) y menos del 50% de los ítems CONSORT en el 28% de la muestra (32/113). La Figura 21 muestra el porcentaje de cumplimiento de los ítems de la guía CONSORT según el tipo de estudio.

Los ítems más bien reportados (en más del 85% de los estudios) en todos los tipos de estudios fueron 1b (resumen estructurado), 2a (antecedentes), 2b (objetivos), 4a (criterios de elegibilidad), 6a (resultados), 12a (métodos estadísticos), 14b (por qué se detuvo el estudio), 15 (tabla de datos iniciales). Los EPs reportaron más elementos en más del 85% de los estudios (generación de secuencia 8a, flujo de participantes 13a, 16 números analizados y 17a tamaño del efecto) en comparación con los estudios de efectividad no autodenominados como pragmáticos.

Aunque la descripción de la intervención para cada grupo de estudio fue buena en más del 80% (95/113) de los estudios, los autores no reportaron suficientes detalles de la

85

intervención, la estandarización o individualización de la intervención y las fuentes adicionales para parecerse a la práctica clínica. El ítem 5b (Descripción de los diferentes componentes de las intervenciones y, cuando corresponda, descripción del procedimiento para adaptar las intervenciones a los participantes individuales) se informó mejor en el grupo EE placebo (77% de los estudios) que en el EEotrainintervención (56%) y los EPs (54%). Además, menos del 11% (12/113) de los estudios reportaron si se evaluó la adherencia de los participantes o de los proveedores de la intervención y cómo (ítem 5e).

Los elementos 7a (cómo se determinó el tamaño de la muestra), 11a (enmascaramiento) y 13a (flujo de participantes) se reportaron peor en EEotrainintervención (46%/56%/67%) que en EE placebo (66%/83%/83%) y EP. (74%/77%/87%). Todos los elementos relacionados con el enmascaramiento se reportaron mejor cuando se utilizó un grupo de control con placebo. Los elementos relacionados con la generación de la ocultación de la asignación de la secuencia, la implementación y la generalización de los resultados del estudio se informaron mejor en los EPs que en los estudios de efectividad no autodenominados pragmáticos.

Además, los elementos "efectos adversos" (elemento 19) y "número de registro del estudio y dónde se puede evaluar el protocolo" (elementos 23 y 24 de la guía CONSORT) no se informaron en menos del 50% (55/113) del total de la muestra.

4.3.5 Evaluación del Riesgo de sesgo

El sesgo de "realización" estuvo presente en todos los tipos de estudio. El sesgo de selección (generación de secuencia aleatoria y ocultamiento de la asignación) fue menor en los EPs que en los EE ($p = 0,031$; $p = 0,041$). El sesgo de detección fue mayor en los EPs y los EEotrainintervención que en los EE placebo ($p = 0,012$).

Los resultados mostraron que el porcentaje de cumplimiento de los ítems CONSORT fue

mayor en los ensayos con bajo RoB en los siguientes dominios: generación de secuencias aleatorias, ocultación de la asignación, datos de resultados incompletos e informes selectivos.

4.3.6. Tercera publicación

Roura, S., Alvarez, G., Hohenschurz-Schmidt, D., Solà, I., Núñez-Cortés, R., Bracchiglione, J., Fernández-Jané, C., Phalip, J., Gich, I., Sitjà-Rabert, M., & Urrútia, G. (2025). **Alignment between the research question, design and terminology is required in manual therapy trials - a methodological study.** *Journal of Clinical Epidemiology*, 183, 111817.

ORIGINAL RESEARCH

Alignment between the research question, design, and terminology is required in manual therapy trials: a methodological study

S. Roura^a, G. Alvarez^{b,c,*}, D. Hohenschurz-Schmidt^d, I. Solà^{b,e,f}, R. Núñez-Cortés^g,
J. Bracchiglione^{b,c,h}, C. Fernández-Jané^{c,i,j}, J. Phalip^{k,l}, I. Gich^{c,m}, M. Sitjà-Rabert^{c,i},
G. Urrútia^{b,c,f}

^aPhD Student in Biomedical Research Methodology and Public Health in the Medical Department of the Universitat Autònoma de Barcelona, Barcelona, Spain

^bIberoamerican Cochrane Centre, Institut de Recerca Sant Pau (IR SANT PAU), Barcelona, Spain

^cFacultat de Ciències de la Salut Blanquerna, Universitat Ramon Llull, Barcelona, Spain

^dImperial Clinical Trials Unit, School of Public Health, Imperial College, London, United Kingdom

^eCentro de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP), Madrid, Spain

^fUniversitat Autònoma de Barcelona, Bellaterra, Spain

^gFaculty of Medicine, Department of Physical Therapy, University of Chile, Santiago, Chile

^hInterdisciplinary Centre for Health Studies (CIESAL), Universidad de Valparaíso, Viña del Mar, Chile

ⁱGlobal Research on Wellbeing (GRoW) Research Group, Universitat Ramon Llull, Barcelona, Spain

^jTecnocampus, Universitat Pompeu Fabra, Mataró-Maresme, Barcelona, Spain

^kFaculté de médecine, Fondation ANALGESIA, Clermont-Ferrand, France

^lCHU Clermont-Ferrand, Inserm 1107 Neuro-Dol, Service de pharmacologie médicale, Université Clermont Auvergne, Clermont-Ferrand, France

^mInstitut de Recerca Sant Pau (IR SANT PAU), Barcelona, Spain

Accepted 30 April 2025; Published online 8 May 2025

Abstract

Objectives: The study aimed to assess the design, reporting, and risk of bias in effectiveness trials (ETs) in manual therapy (MT), comparing pragmatic with nonpragmatic trials and trials with and without placebo controls.

Study Design and Setting: We searched MEDLINE and the Cochrane Central Register of Controlled Trials for randomized controlled trials with the term “effectiveness” in the title or abstract in the field of MT from inception to January 2024. Two independent reviewers extracted data on specific study characteristics, their reporting, and risk of bias and assessed them using the PRagmatic Explanatory Continuum Indicatory Summary-2 tool (PRECIS-2). Descriptive analysis using frequencies and percentages and a relation analysis between PRECIS-2 scores and specific study characteristics were performed.

Results: Of the 113 trials, 39 were self-labeled as pragmatic, 39 used usual care or other interventions, and 35 used placebo controls. ETs have increased in recent years. They are moderately pragmatic, whether they are self-described as pragmatic or not, and whether they use a placebo control group or not. Despite their aim to resemble clinical practice, the pragmatic features of these trials are often unclear. Pragmatic features are common in trials’ eligibility, recruitment, and outcome domains, but intervention and setting are rated as very explanatory. Compared to self-labeled pragmatic trials, “effectiveness” trials are less likely to follow reporting guidelines.

Conclusion: The term effectiveness is used in MT trials independently of its pragmatic connotations. Using a placebo or active control does not modify the pragmatic attitude of the ETs. © 2025 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

Keywords: Effectiveness clinical trials; Musculoskeletal manipulations; Randomized controlled trials; Manual therapy; Generalizability; PRECIS-2

1. Introduction

Both efficacy and effectiveness randomized clinical trials (RCTs) are essential for evaluating interventions [1,2]. Efficacy trials are conducted under controlled conditions to explore the

* Corresponding author. Iberoamerican Cochrane Center, C. Sant Antoni Maria Claret 167, Pavelló 18, Planta 008025, Catalunya, Barcelona, Spain.

E-mail address: gerardalv@gmail.com (G. Alvarez).

<https://doi.org/10.1016/j.jclinepi.2025.111817>

0895-4356/© 2025 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

What is new?**Key findings**

- An increasing proportion of trials in manual therapy (MT) are aimed at evaluating effectiveness, as a primary aim.
- ETs in MT have a moderately pragmatic attitude regardless of whether they self-label themselves as pragmatic or not and whether they use a placebo control group or not. Most ETs aim to resemble clinical practice, but pragmatic features are not shown within the trial.
- Reporting ETs is often inadequate, exacerbated when studies do not call themselves pragmatic. It is also poorer when “usual care” or other interventions are used as a control intervention rather than a placebo control.
- Although selection bias is lower when studies call themselves pragmatic, detection bias is lower when using a placebo control than another intervention or “usual care”.

What this adds to what is known?

- The external validity of ETs is often compromised.
- The use of the term efficacy/effectiveness seems unclear among the MT research community, which uses the term effectiveness in efficacy designs.

What is the implication and what should change now?

- A thorough justification for the trial’s aim, rationale, and design can improve study quality.
- Modifications to trial methodologies are essential to enhance their clinical applicability in the MT field.
- Clear and comprehensive reporting is vital for understanding trial quality, mitigating biases, and improving the applicability of findings for clinical practice.

causal relationship between interventions and their physiological effects (explanatory attitude) [3]. Conversely, effectiveness trials (ETs) aim to inform health care professionals by comparing interventions under real-world conditions (pragmatic attitude) [3,4]. The terms commonly used in the literature to designate an article as explanatory are efficacy trials and “pragmatic”, or “naturalistic”, to designate them as pragmatic, or referring to the “effectiveness” of an intervention [3,5–9]. These trial types are better understood as a continuum rather than a strict dichotomy, as they

often share overlapping methodologies and conceptual frameworks [4,10–13]. Furthermore, the terms “attitude” and “study intent” must be differentiated. The chosen “attitude” is the overarching orientation of the research, either choosing a more pragmatic or a more efficacy-focused orientation. In contrast, the “study intent” is the goal pursued within that orientation, in other words, how the author states and justifies the specific objective or research question [3,10], such as “determining if a specific manual therapy (MT) technique is correlated with a specific heart rate variability change (explanatory attitude)” vs “evaluating if adding MT to the standard treatment for obstructive pulmonary infection outperforms the standard treatment alone in real-world conditions (pragmatic attitude)”. ETs are necessary because they focus on increasing research’s relevance and applicability to clinical practice and policy decision-making. Thus, they might ensure external validity but also use methods that enhance internal trial validity, such as randomization and treatment fidelity monitoring.

In MT, efficacy and ETs are well-documented [1,14]. However, concerns have been raised regarding their overall methodological rigor due to internal validity issues and reporting deficiencies [14–22]. Although the term “effectiveness” is commonly used in MT trials [23–26], a systematic review revealed a lack of pragmatism in many cases [1]. This has led to concerns about whether these trials accurately simulate clinical practice and deliver meaningful insights for health care decisions [1,27,28]. The inconsistent use of terminology has implications for research, potentially affecting the interpretation of trial outcomes and influencing clinical guideline recommendations [29,30].

Research shows that MT trials often lean toward explanatory designs rather than pragmatic approaches [1]. Previous findings from the author group confirmed that self-labeled MT pragmatic RCTs (note that to facilitate readability, those will be named pragmatic trials (PTs) throughout the manuscript) have not increased in recent years, unlike in other disciplines [1,10,31–33]. Furthermore, their pragmatic attitude remains moderate, and better internal validity and reporting are needed [33]. The misuse of “efficacy” and “effectiveness” as interchangeable terms has resulted in research designs often inconsistent with their stated objectives [4,30,34,35]. This issue is particularly prevalent in studies focused on the clinical effects of interventions with limited efficacy evidence but widespread clinical use, such as MT [31,36].

Given the conceptual uncertainty regarding terminology, this study aimed to compare effectiveness MT trials self-labeled as pragmatic with those that are not. A substantial body of research exists that neither applies well in real-world settings nor is designed with the internal validity required to be considered efficacy studies. By assessing the methodological differences, the researchers sought to clarify whether there are critical design disparities and promote more accurate reporting and understanding of trial

characteristics. This research underscores the need for trial design and terminology clarity to strengthen academic inquiry and clinical guideline development. Establishing consistent definitions and methodological practices will help improve the quality and reliability of MT trials, ultimately benefiting clinical practice and health care decision-making.

2. Methods

2.1. Protocol registration

We conducted a methodological review of ETs on MT interventions, following the guidelines of the Cochrane Handbook for systematic reviews of interventions [37] and reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines (PRISMA) [38] ([Supplementary file 1](#)). The study was prospectively registered on the Open Science Framework (DOI 10.17605/OSF.IO/WKEPZ).

2.2. Eligibility criteria

We included published RCTs with the terms “effectiveness” (related to an intervention), “pragmatic” (methodological design), or “naturalistic” (methodological design) in either the title or abstract [6,7]. Eligible references had to include either a manual technique or a combination of manual techniques (soft tissue techniques, joint mobilization or manipulation, massage, myofascial release, nerve manipulation, strain/counterstrain, and acupressure). There were no restrictions regarding population, comparator groups or outcome measures. Exclusion criteria included experimental interventions using tools, devices (electrotherapy, kinesio taping, dry needling, and acupuncture), drugs, active exercises, or a combination of therapies without MT intervention. We also excluded non-English articles, protocols, and poster/conference presentations.

2.3. Search strategy

We conducted a comprehensive search of MEDLINE and the Cochrane Central Register of Controlled Trials from inception to January 2024, using a search strategy that combined controlled vocabulary with relevant Medical Subject Headings terms in the field of MT and the design of interest (see [Supplementary file 2](#)). The strategy was developed by an expert methodologist on our research team (L.S.).

2.4. Study selection

The records were uploaded onto the Rayyan software (www.rayyan.ai) [39]. On deduplication, the references were screened by two independent reviewers (S.R. and G.A.) based on the title and summary, resolving disagreements through discussion.

The references were initially classified into two groups: group 1 consisting of ETs labeled as “pragmatic” in the title and/or abstract (PTs), and group 2 comprising ETs that were not self-labeled as pragmatic. For further analysis, we split the ET group into studies that used a placebo/sham control group (effectiveness Randomised Controlled Trial using a placebo/sham comparator [ETpl])—GROUP 2A—and those that used an accepted or established comparator (effectiveness Randomised Controlled Trial using an active comparator [ETac])—GROUP 2B—based on title/abstract information, resulting in three groups (groups 1, 2A, and 2B). We determined the sample of studies to be included in each group according to the PTs found. As we found many more ETs than PTs, we randomized the ETs by matching their number to the number of PTs found.

2.5. Data collection process

Two reviewers, S.R. and an additional member from G.A., R.N., J.B., D.H., C.F., and J.P., gathered data from the included studies and resolved disagreements through consensus. The involvement of a third party was not required to resolve any discrepancies.

Before the review process, the team met with Dr Kirsty Loudon, the developer of the pragmatic-explanatory continuum indicator summary (PRECIS)-2 tool [40]. It was deemed suitable to involve a member of the PRECIS-2 tool authors, given that PRECIS-2 was conceptualized as a tool for trial designers rather than as a method for retrospective assessment of pragmatism. The objective was to ensure that the method was used in the most reliable manner. Three articles were examined, compared, and discussed against Dr Loudon’s criteria.

2.6. Data items

Three authors (S.R., G.A., and G.U.) designed a data extraction form. It included bibliometric identification elements, the intent of the trial, the rationale of the intervention given by the authors, the experimental and control interventions, the limitations reported by the authors, the PRECIS-2 tool assessment, the Consolidated Standards of Reporting Trials (CONSORT) reporting assessment, and the risk of bias (RoB) assessment. Information about the data extraction form and the methods used to assess each item are included in [Supplementary file 3](#) and reported elsewhere [33].

The data extraction form was piloted to assess reviewers’ discrepancies. Subsequently, all the suggestions were incorporated. Finally, a guideline for reviewers was provided to the team ([Supplementary file 3](#)). When information about a PRECIS-2 domain was not found within the trial publication, it was rated as “blank”

2.7. Data analysis

A descriptive analysis of the categorical variables was conducted, wherein results were presented as relative and

absolute frequencies. Likewise, means and SDs were used to describe quantitative variables. Median and IQRs were applied for ordinal variables. Meanwhile, the PRECIS-2 wheel was used to present the pragmatic attitude of each article graphically. Furthermore, variables were compared using analysis of variance for quantitative variables, Chi-square for categorical variables and Kruskal–Wallis for ordinal variables. According to previous research [41–43], we used a 9–45 point scale to evaluate the total score PRECIS-2, apart from the PRECIS average. Ratings of 9–22 were described as slightly pragmatic, 23–34 moderately pragmatic, and ≥ 35 very pragmatic. We conducted post hoc analysis using the Scheffé test for analysis of variance and the Mann–Whitney test for Kruskal–Wallis. The Spearman test was performed to analyze the correlation between PRECIS-2 and some descriptive characteristics. The significance level was set at 5%

($\alpha = 0.05$). All data were analyzed using IBM-SPSS software (V26.0).

3. Results

3.1. Study selection

The search identified 3553 articles. After screening the title, abstract, and full-text availability, the final sample comprised 306 studies. Among these, 39 were identified as self-labeled pragmatic in the title and/or abstract, whereas 267 were ETs not self-labeled as pragmatic. We did not identify any articles in MT with the term naturalistic. Of the 267 ETs, 35 were ETpl, whereas the remaining 232 used other accepted or established interventions (ETac). To balance and compare the three groups, the 232 ETac were randomized using the Microsoft Excel random generator to get a group of

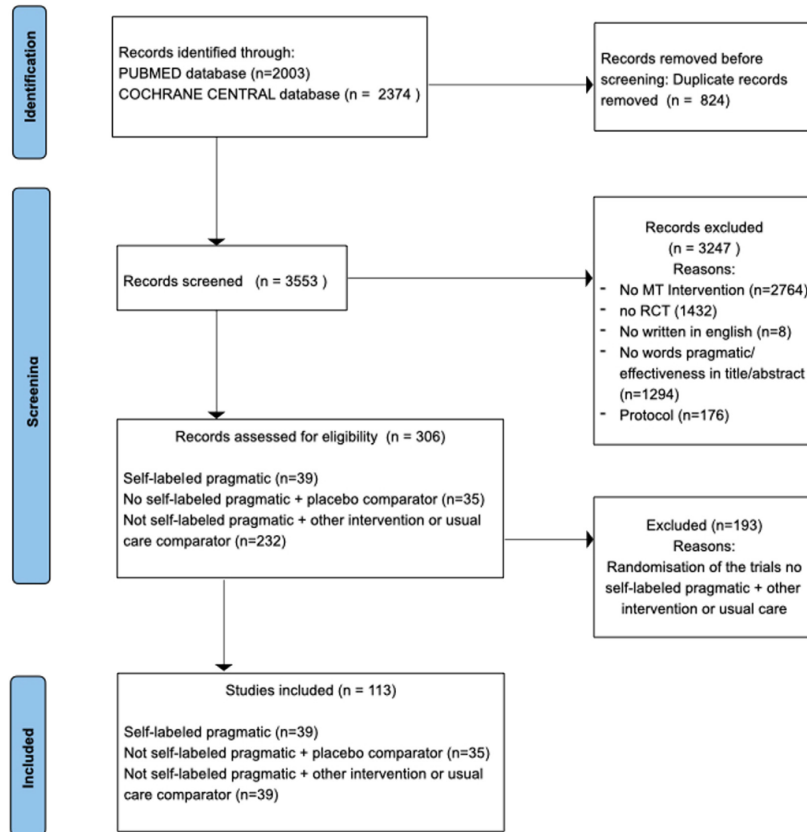


Figure 1. PRISMA flow diagram. MT manual therapy, PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses, RCT, randomized controlled trial.

39 trials (the same number of trials as in the PTs group), resulting in a total sample for this review of 113 studies. The PRISMA diagram is shown in [Figure 1](#).

3.2. Study characteristics of the sample

[Table 1](#) summarizes the main characteristics of our sample. A detailed table with all included studies and some additional characteristics can be found in [Supplementary File 4](#). The average number of participants was significantly higher in PTs (mean = 169) compared to ETac (mean = 65) and ETpl (mean = 70). In PTs, the most frequent intervention studied was a nonprotocolized combination of techniques (54%; 21/39), whereas in ETac, it was a protocol of a combination of techniques (51%; 20/39) and a protocol of an isolated technique in 49% (17/35) of the ETpl. No studies blinded therapists. Blinding of external assessors was more common in ETpl (74%) compared to PTs (69%) and ETac (59%). Multicentric settings were more prevalent in PTs (56%), whereas ETs primarily used unicentric settings.

Specific characteristics of PTs were reported elsewhere [33,44]. Although “pragmatic” was not mentioned in titles or abstracts, 64% of ETac studies claimed a pragmatic approach, compared to 31% of ETpl. Common self-reported study limitations included participant blinding (14/17/12), generalizability (13/11/11), and sample size (11/12/14). Inadequate control was frequent in PTs (13) and ETac (8) but rare in ETpl (4). Study length limitations were prevalent in ETs. [Figure 2](#) illustrates these limitations across study types.

3.3. PRECIS assessment

[Table 2](#) presents mean PRECIS-2 scores, revealing that 60% of studies in each group had a moderately pragmatic attitude. [Figure 3](#) illustrates the PRECIS-2 wheel based on study group.

[Table 3](#) shows PRECIS-2 rates by study type. Most domains in PTs, except domain 6 (adherence), were pragmatic in over 50% of studies. For ETs, domains 1 (eligibility), 2 (recruitment), and 8 (outcome measure) were pragmatic in more than 50% of cases. Domain 3 (setting) was largely explanatory in ETs (82%), whereas domain 5 (flexibility) was pragmatic in over 50% of PTs but explanatory in most ETs.

3.4. CONSORT assessment

Almost one-third of the trials (37/113) were reported according to the CONSORT statement: 41% (16/39) of the PTs, 31% (12/39) of the ETac, and 26% (12/35) of the ETpl.

One in four trials reported details related to more than 70% of CONSORT items (13/39 of the PTs, 9/39 of the ETac, and 13/32 of the ETpl), between 50% and 70% in 41% of the trials (46/113) and lower than 50% of the

CONSORT items in 28% of the sample (32/113). [Figure 4](#) shows the percentage of compliance with CONSORT items depending on the study type.

[Supplementary file 5](#) shows item compliance across study types. The data analysis from the pre-CONSORT and post-CONSORT periods indicates that the reporting of some items improved after the publication of the CONSORT tool. However, this improvement does not impact each item’s final analyzed reporting result ([Supplementary file 6](#)). The items better reported (more than 85% of the trials) in all study types were structured summary, background, objectives, eligibility criteria, outcomes, statistical methods, why the trial stopped, and baseline data table. Pragmatic RCTs had more items reported in more than 85% of the trials (sequence generation, participants flow, numbers analyzed, and effect size) compared with ETs that were not self-labeled pragmatic.

Intervention descriptions were adequate in over 80% of studies (95/113). Still, the authors did not sufficiently report details of the intervention, the standardization or individualization of the intervention and the additional sources to resemble clinical practice. Item 5 b (Description of the different components of the interventions and, when applicable, description of the procedure for tailoring the interventions to individual participants) was better reported in the ETpl group (77%) than the ETac (56%) and the PTs groups (54%). Moreover, less than 11% (12/113) of the trials reported whether and how the adherence of participants or care providers was assessed.

How sample size was determined, blinding, and participants’ flow were reported less optimally in ETac (46%, 56%, 67%, respectively), compared to ETpl (66%, 83%, 83%, respectively) and PTs (74%, 77%, 87%, respectively) groups. All the items regarding blinding were better reported when using a placebo/sham control group. Items regarding randomization sequence generation, allocation concealment, implementation, and generalizability of the trial results were reported better in PTs than in ETs.

Furthermore, items “harms” and “trial registration number”, and “where the protocol can be assessed” were not reported in the 49% (55/113) of the sample.

3.5. Risk of bias

Selection bias (random sequence generation and allocation concealment) was lower in PTs than in ETs. Detection bias was higher in PTs and ETac than ETpl. A table comparing the RoB between types of studies is presented in [Supplementary file 7](#).

Please refer to [Supplementary file 8](#) to compare the studies with a low vs high/unclear RoB according to compliance with CONSORT items. The results showed that the percentage of compliance with CONSORT items was higher in the trials with low RoB in the following domains: random sequence generation, allocation

Table 1. Main characteristics of the sample

Characteristics of the sample	PTs (N = 39)	ETac (N = 39)	ETpl (N = 35)
Number of participants (Mean)	169 (SD = 151)	65 (SD = 58)	70 (SD = 44)
Number of participants	Percent (n ^b)		
Intervention			
Combination of nonprotocolized techniques	54% (21)	10% (4)	9% (3)
Protocol of a combination of techniques	13% (5)	51% (20)	23% (8)
Isolated nonprotocolized technique	15% (6)	8% (3)	8% (3)
Protocol of an isolated technique	5% (2)	18% (7)	49% (17)
Combination of nonprotocolized therapies	10% (4)	3% (1)	3% (1)
Protocol of a combination of therapies	3% (1)	10% (4)	9% (3)
Control intervention			
2 arms			
Test treatment vs other active intervention	31% (12)	46% (18)	0% (0)
Test treatment vs placebo	3% (1)	0% (0)	69% (24)
Test treatment vs usual care	33% (13)	38% (15)	0% (0)
Test treatment vs no intervention	13% (5)	5% (2)	0% (0)
3 arms			
Test treatment vs 2 other active interventions	5% (2)	3% (1)	0% (0)
Test treatment vs 1 other active intervention and 1 placebo	2% (1)	3% (1)	11% (4)
Test treatment vs 1 active intervention and 1 usual care	10% (4)	3% (1)	0% (0)
Test treatment vs 1 usual care and 1 placebo	2% (1)	0% (0)	17% (6)
Test treatment vs 1 other active intervention and 1 no intervention	2% (1)	0% (0)	0% (0)
4 arms			
Test treatment vs 1 other active intervention, 1 no intervention and 1 placebo	0% (0)	3% (1)	3% (1)
Blinding (yes)			
Participants	8% (3)	8% (3)	20% (7)
Therapists	0% (0)	0% (0)	0% (0)
External assessors to providers and patients	69% (27)	59% (23)	74% (26)
Statistician	43% (17)	18% (7)	20% (7)
Rationale			
Yes (rationale provided)	64% (25)	64% (25)	63% (22)
Comparative effectiveness data	61%	38%	31%
Mechanistic experiments	36%	54%	63%
Highly sham-controlled	3%	8%	6%
No (no rationale provided)	21% (8)	28% (11)	26% (9)
Unclear	13% (5)	8% (3)	11% (4)
Funding			
Yes	67% (26)	21% (8)	37% (13)
Follow-up			
No follow-up	18% (7)	23% (9)	28% (9)
<2 wk	0% (0)	13% (5)	3% (1)

(Continued)

Table 1. Continued

Characteristics of the sample	PTs (N = 39)	ETac (N = 39)	ETpl (N = 35)
2–4 wk	2% (1)	8% (3)	3% (1)
4–12 wk	10% (4)	20% (8)	34% (12)
3–6 mo	18% (7)	15% (6)	20% (7)
6–12 mo	20% (8)	15% (6)	9% (3)
> 1 y	26% (10)	5% (2)	3% (1)
Individualized ^a	5% (2)	0% (0)	0% (0)
Setting			
Multicentric	56% (22)	8% (3)	20% (7)
Unicentric	33% (13)	87% (34)	77% (27)
Unclear reported	10% (4)	5% (2)	3% (1)

ETac, effectiveness trials active comparator; ETpl, effectiveness trials placebo comparator; PTs, pragmatic trials.

^a Trials assessing how many weeks patients remained pain-free.

^b Number of studies = 113.

concealment, incomplete outcome data, and selective reporting.

4. Discussion

This review compared the design, reporting, and bias risk of MT effectiveness RCTs, comparing trials self-labeled as pragmatic or not by the study authors, and using a placebo control or an active comparator. The results

showed a moderate pragmatic attitude regardless of labeling or placebo use. Although many trials aim to emulate clinical practice, designs often lack pragmatic methodology. The review also found that reporting was generally inadequate, especially when not self-labeled pragmatic or when using usual care/other interventions instead of placebo/sham controls as comparators. The findings highlight areas for improvement in designing and reporting MT effectiveness RCTs to better align with clinical practice and enhance methodological rigor.

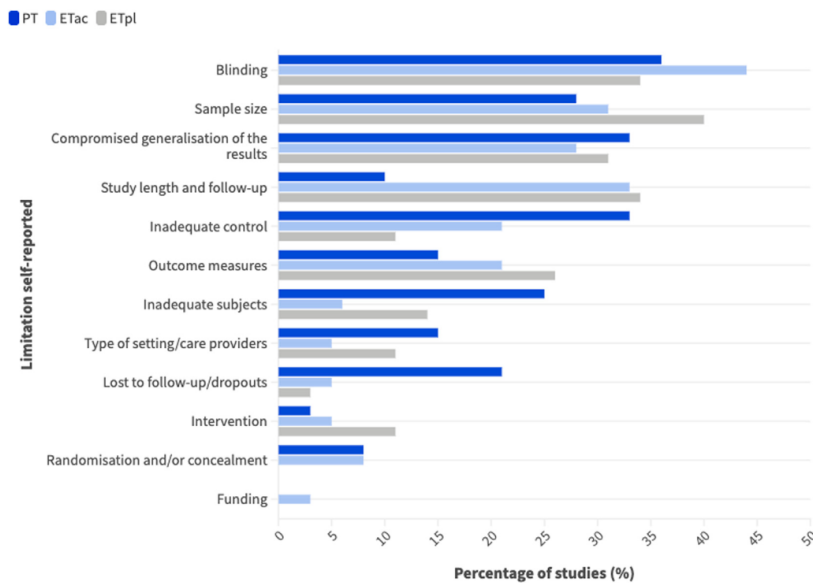


Figure 2. Comparison of limitations reported between studies. ETac, effectiveness trials active comparator; ETpl, effectiveness trials placebo comparator; PTs, pragmatic trials.

Table 2. Mean PRECIS-2 scores between studies and pragmatic attitude

Type of study	Mean PRECIS score (SD)	Slightly pragmatic	Moderate pragmatic	Highly pragmatic
PTs	3,5 (0,6)	3% (1)	61% (24)	36% (14)
ETac	2,8 (0,6)	33% (13)	59% (23)	8% (3)
ETpl	2,8 (0,6)	34% (12)	63% (22)	3% (1)

ETac, effectiveness trials active comparator; ETpl, effectiveness trials placebo/sham comparator; PTs, pragmatic trials.

4.1. Designing MT effectiveness RCTs and correct use of the terminology

Although the growing interest in ETs designed to reflect clinical practice, probably a response to the demand for real-world assessments of interventions, the distinction between efficacy and ETs designs remains unclear [4]. Two key factors define a trial's position on the efficacy-effectiveness continuum: the study's intention [10] and the rationale for its design [4]. The intention determines the research question, affects the design, and influences the outcomes and applicability of the study [4,10]. When giving a study's rationale, authors should determine if sufficient efficacy trials justify evaluation in more realistic environments [4]. Our analysis revealed that 25% of trials lacked this rationale, and those that did often relied on mechanistic experiments rather than controlled placebo/sham trials. As previous studies suggest, our findings reveal a misalignment between trial objectives and methodologies, warranting further investigation into better design practices [1].

4.2. Pragmatic attitude and methodological characteristics of effectiveness MT RCTs

Effectiveness and PTs are terms that can be used interchangeably in many contexts to refer to clinical trials designed to answer whether an intervention works under routine health care conditions [9]. However, ETs scored lower on PRECIS-2 (2.8) than self-labeled PTs (3.5),

suggesting a greater emphasis on pragmatic design when explicitly labeled. In our sample, in contrast to PTs, ETs were primarily explanatory in the setting and intervention domains and had smaller sample sizes.

4.2.1. Experimental intervention

The most common interventions in ETs were protocolized combinations of techniques (ETac) and isolated techniques (ETpl), placing these studies closer to the efficacy pole of the continuum. Adopting nonprotocolized, patient-centered interventions could better reflect clinical practice [36,45,46]. In nonpharmacological research, such approaches are essential for generating relevant and actionable clinical results [1]. However, the lack of standardized interventions poses a significant challenge. Guidelines such as Template for Intervention Description and Replication [47] are essential for effective reporting and replicability, including detailed annexes to avoid interpretative gaps.

In addition, the use of placebo/sham comparators, uncentric settings, and blinding, which is common in explanatory studies, has led to a debate on their pragmatism [45–49]. Whether the same discussion should pertain to nonlabeled pragmatic ETs is debatable, as “effectiveness” is often deemed synonymous with “pragmatic” in research contexts [5–8].

4.2.2. Placebo/sham control

Including placebo/sham control interventions in MT RCTs may clarify therapeutic effects, which are

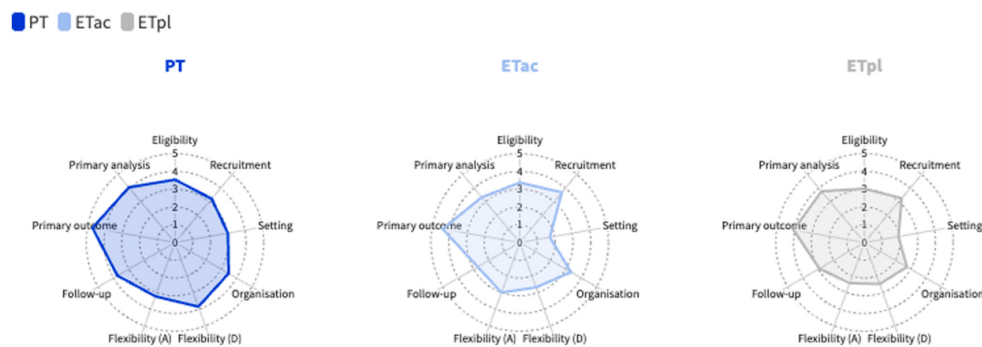


Figure 3. PRECIS-2 wheels of each study type. ETac, effectiveness trials active comparator; ETpl, effectiveness trials placebo/sham comparator; PRECIS, PRagmatic EXplanatory Continuum Indicator Summary; PTs, pragmatic trials.

Table 3. PRECIS-2 rates for each type of study

Type of study	PRECIS-2 score	Eligibility	Recruitment	Setting	Organization	Flexibility intervention	Flexibility adherence	Follow-up	Primary outcome	Primary analysis
PTs (n = 39)										
	4/5	54% (21)	54% (20)	55% (21)	51% (20)	65% (25)	39% (12)	62% (24)	95% (37)	71% (27)
	1/2	13% (5)	35% (13)	39% (15)	23% (9)	8% (3)	19% (6)	10% (4)	3% (1)	21% (8)
	3	33% (13)	11% (4)	5% (2)	26% (10)	28% (11)	42% (13)	28% (11)	3% (1)	8% (3)
	Blank	0% (0)	5% (2)	3% (1)	0% (0)	0% (0)	26% (8)	0% (0)	0% (0)	3% (1)
ETac (n = 39)										
	4/5	51% (20)	56% (22)	5% (2)	41% (16)	18% (7)	15% (6)	15% (6)	92% (36)	41% (16)
	1/2	28% (11)	18% (7)	82% (32)	15% (6)	51% (20)	15% (6)	41% (16)	3% (1)	28% (11)
	3	18% (7)	8% (3)	5% (2)	33% (13)	31% (12)	23% (9)	44% (17)	5% (2)	15% (6)
	Blank	3% (1)	18% (7)	8% (3)	10% (4)	0% (0)	46% (18)	0% (0)	0% (0)	15% (6)
ETpl (n = 35)										
	4/5	31% (11)	46% (16)	20% (7)	23% (8)	11% (4)	6% (2)	17% (6)	83% (29)	74% (26)
	1/2	31% (11)	31% (11)	74% (26)	43% (15)	60% (21)	34% (12)	23% (8)	9% (3)	26% (9)
	3	34% (12)	6% (4)	3% (1)	31% (11)	26% (10)	31% (11)	60% (21)	9% (3)	0% (0)
	Blank	3% (1)	11% (4)	3% (1)	3% (1)	0% (0)	29% (10)	0% (0)	0% (0)	0% (0)

ETac, effectiveness trials active comparator; ETpl, effectiveness trials placebo/sham comparator; PRECIS, PRagmatic EXplanatory Continuum Indicatory Summary; PTs, pragmatic trials.

Blank: percentage of studies without information about the domain within the study.

challenging to isolate in efficacy studies. The ongoing PRECIS-3 update aims to assess control groups more effectively, highlighting the need for further research on the role of placebos/shams in PTs, particularly in MT studies where efficacy trials struggle to pinpoint mechanisms despite demonstrating clinical outcomes.

Nonpharmacological interventions differ from pharmacological ones in development and implementation, often lacking prior efficacy studies. Our review found no significant variations in the PRECIS-2 scores between trials using placebo/sham or other comparators, indicating a moderately pragmatic approach. This finding is consistent

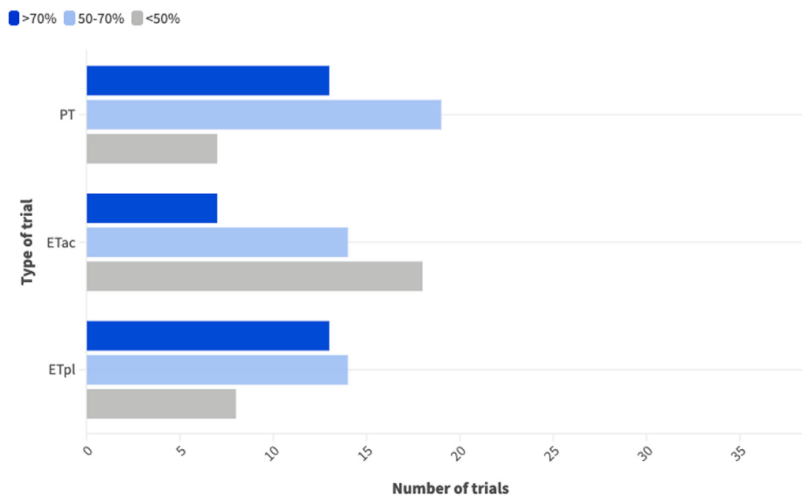


Figure 4. Compliance with CONSORT items. CONSORT, Consolidated Standards of Reporting Trials; ETac, effectiveness trials active comparator; ETpl, effectiveness trials placebo/sham comparator; PTs, pragmatic trials.

with the analysis conducted by Devos et al on nursing interventions [41]. Dal-Re et al argue that placebo-controlled trials cannot replicate clinical practice and should not be considered pragmatic. However, the PRECIS-2 authors believe that placebo/sham trials can still inform clinical decisions [12,48–52]. Dal-Re et al also found PRECIS-2 insufficiently sensitive to differentiate between masked and open-label trials [50]. The placebo effects in “touch” interventions are considered part of MT treatments’ therapeutic outcomes [53,54]. Although high-quality efficacy trials are required to clarify whether a MT intervention is more than a placebo/sham [55], including a placebo/sham control in MT RCTs (although it has explanatory connotations) could aid in understanding the real effect of MT interventions that are challenging to achieve in an efficacy study under less realistic conditions. The ongoing PRECIS-3 update aims to assess control groups more effectively, highlighting the need for further research on the role of placebos in PTs, which will be particularly interesting in MT [50–52,56].

4.2.3. Setting

Our results showed that many ETs were conducted in a single center, limiting results’ generalizability [48]. However, Zwarenstein et al argue that detailed reporting of the environment, health system, and patient types can still provide valuable clinical insights [49]. Although multi-centre trials are preferable for broader applicability, they are costly and challenging for MT researchers, who typically lack adequate public and private funding, a significant barrier to more extensive research.

4.2.4. Blinding

PRECIS-2 authors acknowledge that although blinding poses challenges, it remains beneficial in pragmatic studies by significantly enhancing decision-making data quality through the control of the placebo effect and subjectivity [49,56]. In MT, where mechanistic evidence and placebo-controlled studies are scarce, maintaining blinding in trials might help to preserve the rigor of RCTs.

4.3. Reporting and quality of effectiveness MT RCTs

Accurate trial reporting is essential for the applicability and quality of results [4,21]. Our review found that ETac were reported less comprehensively than PTs and ETpl. It is crucial to include adequate details on trial settings, participants, and interventions to facilitate external implementation and validation [4]. In MT, biases are common due to challenges in blinding therapists and patients, especially when patient-reported outcomes are used. This supports a pragmatic approach but poses a high RoB, reducing internal validity.

Our findings indicate that the external validity of ETs is often compromised. A thorough justification for the trial’s

aim, rationale, and design can improve study quality and interpretation. Modifications to trial methodologies are essential to enhance their clinical applicability in the MT field. Clear and comprehensive reporting is vital for understanding trial quality, mitigating biases, and improving the applicability of findings for clinical practice.

4.3. Strengths and limitations

Although the small number of studies in each review group might limit the generalizability of our findings, we intentionally split the sample to compare “effectiveness” studies with self-labeled PTs—ensuring that the groups were matched by the number of PTs. In addition, we acknowledge that not all MTs may be equally represented and that the inclusion of diverse reviewers could have introduced bias into the PRECIS-2 scores. To mitigate subjectivity, we implemented extensive training, thorough piloting of the data, and comprehensive discussions with the principal investigator for precise data extraction. We also recognize that using trial design as a proxy for the underlying intent is a limitation; ideally, direct consultation with the research teams would yield a more accurate assessment. However, this approach highlights the importance of researchers explicitly stating the primary objective of their trials in the published reports. Furthermore, although we acknowledge the limitations of retrospectively applying PRECIS-2, including the caution from its original developers against simply summing the domain scores, we performed a comprehensive evaluation of pragmatism through a meticulous analysis of the PRECIS-2 scores, the authors’ rationale, and the summary score. Our decision was motivated by the need for an objective, systematic comparison of study designs despite the known limitations of the tool.

5. Conclusions

In MT trials, “effectiveness” is used regardless of its pragmatic connotations. These trials are pragmatic regarding participant eligibility, recruitment, and outcome measures but explanatory regarding intervention and setting. The control type (placebo/sham or active) does not affect pragmatism. Compared to self-labeled PTs, ETs show lower compliance with reporting guidelines items.

CRedit authorship contribution statement

S. Roura: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **G. Alvarez:** Writing – review & editing, Supervision, Methodology, Investigation, Conceptualization. **D. Hohenschurz-Schmidt:** Writing –

review & editing, Investigation. **I. Solà:** Methodology. **R. Núñez-Cortés:** Investigation, Formal analysis. **J. Bracchi-gliane:** Methodology, Investigation. **C. Fernández-Jané:** Investigation. **J. Phalip:** Investigation. **I. Gich:** Formal analysis. **M. Sitjà-Rabert:** Supervision. **G. Urrútia:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Declaration of competing interest

There are no competing interests for any authors.

Acknowledgments

Sònia Roura is a PhD student in Biomedical Research Methodology and Public Health in the Medical Department of the Universitat Autònoma de Barcelona, Barcelona, Spain. The authors appreciate the support and dedication of Dr Kirsty Loudon in the early stages of the project, teaching and guiding us with the PRECIS-2 tool. The authors would like to thank Roberto Acosta for his support in the first stages of the project.

Data availability

Data will be made available on request.

References

- [1] Daniel MC, Subialka JA, Young JL, Rhon DI. Over half of clinical trials of mobilization and manipulation for patients with low back pain may have limited real-world applicability. A Systematic Review of 132 Clinical Trials. *J Orthop Sports Phys Ther* 2022;52:532–45.
- [2] Glasziou P, Matthews R, Boutron I, Chalmers I, Armitage P. The differences and overlaps between “explanatory” and “pragmatic” controlled trials: a historical perspective. *J R Soc Med* 2023;116:425–32. <https://doi.org/10.1177/01410768231207536>.
- [3] Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutic trials. *J Chronic Dis* 1967;20:637–48. [https://doi.org/10.1016/0021-9681\(67\)90041-0](https://doi.org/10.1016/0021-9681(67)90041-0).
- [4] Singal AG, Higgins PDR, Waljee AK. A primer on effectiveness and efficacy trials. *Clin Transl Gastroenterol* 2014;5:e45. <https://doi.org/10.1038/ctg.2013.13>.
- [5] Janiaud P, Dal-Ré R, Ioannidis JPA. Assessment of pragmatism in recently published randomized clinical trials. *JAMA Intern Med* 2018;178:1278. <https://doi.org/10.1001/jamainternmed.2018.3321>.
- [6] Dal-Ré R, Janiaud P, Ioannidis JPA. Real-world evidence: how pragmatic are randomized controlled trials labeled as pragmatic? *BMC Med* 2018;16:49. <https://doi.org/10.1186/s12916-018-1038-2>.
- [7] Sacristán JA, Dilla T. Pragmatic trials revisited: applicability is about individualization. *J Clin Epidemiol* 2018;99:164–6. <https://doi.org/10.1016/j.jclinepi.2018.02.003>.
- [8] Ford I, Norrie J. Pragmatic trials. *New Engl J Med* 2016;375:454–63. <https://doi.org/10.1056/nejmr1510059>.
- [9] Fortney JC, Curran GM, Lyon AR, Check DK, Flum DR. Similarities and differences between pragmatic trials and hybrid effectiveness-implementation trials. *J Gen Intern Med* 2024;39:1735–43. <https://doi.org/10.1007/s11606-024-08747-1>.
- [10] Nicholls SG, Zwarenstein M, Hey SP, Giraudeau B, Campbell MK, Taljaard M. The importance of decision intent within descriptions of pragmatic trials. *J Clin Epidemiol* 2020;125:30–7. <https://doi.org/10.1016/j.jclinepi.2020.04.030>.
- [11] Treweek S, Zwarenstein M. Making trials matter: pragmatic and explanatory trials and the problem of applicability. *Trials* 2009;10:37. <https://doi.org/10.1186/1745-6215-10-37>.
- [12] Thorpe KE, Zwarenstein M, Oxman AD, Treweek S, Furlberg CD, Altman DG, et al. A pragmatic–explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *J Clin Epidemiol* 2009;62:464–75. <https://doi.org/10.1016/j.jclinepi.2008.12.011>.
- [13] Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, et al. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *BMJ* 2008;337:a2390. <https://doi.org/10.1136/bmj.a2390>.
- [14] Alvarez G, Solà I, Sitjà-Rabert M, Fort-Vanmeerhaeghe A, Gich I, Fernández C, et al. A methodological review revealed that reporting of trials in manual therapy has not improved over time. *J Clin Epidemiol* 2020;121:32–44. <https://doi.org/10.1016/j.jclinepi.2020.01.006>.
- [15] Riley SP, Swanson BT, Sawyer SF, Brismée J-M. Is research quality in orthopedic manual therapy trials stagnating? Reflections and pathways for improving research quality and advance our profession. *J Man Manip Ther* 2016;24:239–40. <https://doi.org/10.1080/10669817.2016.1253561>.
- [16] Gonzalez GZ, Moseley AM, Maher CG, Nascimento DP, Costa L, da CM, et al. Methodologic quality and statistical reporting of physical therapy randomized controlled trials relevant to musculoskeletal conditions. *Arch Phys Med Rehabil* 2018;99:129–36. <https://doi.org/10.1016/j.apmr.2017.08.485>.
- [17] Koes BW. How to evaluate manual therapy: value and pitfalls of randomized clinical trials. *Man Ther* 2004;9:183–4. <https://doi.org/10.1016/j.math.2004.04.002>.
- [18] Cashin AG, Lee H, Bagg MK, O'Hagan E, Traeger AC, Kanper SJ, et al. A systematic review highlights the need to improve the quality and applicability of trials of physical therapy interventions for low back pain. *J Clin Epidemiol* 2020;126:106–15. <https://doi.org/10.1016/j.jclinepi.2020.06.025>.
- [19] Clar C, Tsertsvadze A, Court R, Hundt GL, Clarke A, Sutcliffe P. Clinical effectiveness of manual therapy for the management of musculoskeletal and non-musculoskeletal conditions: systematic review and update of UK evidence report. *Chiropr Man Therap* 2014;22:12. <https://doi.org/10.1186/2045-709X-22-12>.
- [20] Riley SP, Swanson B, Brismée J-M, Sawyer SF. A systematic review of orthopaedic manual therapy randomized clinical trials quality. *J Man Manipulative Ther* 2016;24:241–52. <https://doi.org/10.1080/10669817.2015.1119372>.
- [21] Innocenti T, Giagio S, Salvio S, Feller D, Minucci S, Brindisino F, et al. Completeness of reporting is suboptimal in randomized controlled trials published in rehabilitation journals, with trials with low risk of bias displaying better reporting: a meta-research study. *Arch Phys Med Rehabil* 2022;103:1839–47. <https://doi.org/10.1016/j.apmr.2022.01.156>.
- [22] Fregni F, Imamura M, Chien HF, Lew HL, Boggio P, Kaptchuk TJ, et al. Challenges and recommendations for placebo controls in randomized trials in physical and rehabilitation medicine: a report of the international placebo symposium working group. *Am J Phys Med Rehabil* 2010;89:160–72. <https://doi.org/10.1097/PHM.0b013e3181bc0bbd>.
- [23] Simoni G, Bozzolan M, Bonini S, Grassi A, Zucchini A, Mazzanti C, et al. Effectiveness of standard cervical physiotherapy plus diaphragm manual therapy on pain in patients with chronic neck pain: a randomized controlled trial. *J Bodyw Mov Ther* 2021;26:481–91. <https://doi.org/10.1016/j.jbmt.2020.12.032>.
- [24] Scafoglieri A, Van den Broeck J, Willems S, Tamminga R, van der Hoeven H, Engelsma Y, et al. Effectiveness of local exercise therapy versus spinal manual therapy in patients with patellofemoral pain syndrome: medium term follow-up results of a randomized controlled trial. *BMC Musculoskelet Disord* 2021;22:446. <https://doi.org/10.1186/s12891-021-04310-9>.
- [25] Muñoz-Gómez E, Inglés M, Serra-Añó P, Espí-López GV. Effectiveness of a manual therapy protocol based on articular techniques in

- migraine patients. A randomized controlled trial. *Musculoskeletal Sci Pract* 2021;54:102386. <https://doi.org/10.1016/j.msksp.2021.102386>.
- [26] Izaola-Azkona L, Vicenzino B, Olabarrieta-Eguia I, Saez M, Lascuain-Aguirrebeña I. Effectiveness of mobilization of the talus and distal fibula in the management of acute lateral ankle sprain. *Phys Ther* 2021;101:pzab111. <https://doi.org/10.1093/ptj/pzab111>.
- [27] Gamerman V, Cai T, Elsaßer A. Pragmatic randomized clinical trials: best practices and statistical guidance. *Health Serv Outcomes Res Methodol* 2019;19:23–35. <https://doi.org/10.1007/s10742-018-0192-5>.
- [28] Wieland L S, Berman B M, Altman D G, Barth J, Bouter L M, D'Adamo C R, et al. Rating of Included Trials on the Efficacy-Effectiveness Spectrum: development of a new tool for systematic reviews. *J Clin Epidemiol* 2017;84:95–104. <https://doi.org/10.1016/j.jclinepi.2017.01.010>.
- [29] Deaton A, Cartwright N. Understanding and misunderstanding randomized controlled trials. *Soc Sci Med* 2018;210:2–21. <https://doi.org/10.1016/j.socscimed.2017.12.005>.
- [30] Porzolt F, Wiedemann F, Philippen M, Weiss C, Weiss M, Schmalting K, et al. The terminology conflict on efficacy and effectiveness in healthcare. *J Comp Eff Res* 2020;9:1171–8. <https://doi.org/10.2217/ceer-2020-0149>.
- [31] Hohenschurz-Schmidt D, Kleykamp B A, Draper-Rodi J, Vollert J, Chan J, Ferguson M, et al. Pragmatic trials of pain therapies: a systematic review of methods. *Pain* 2021;163:21–46.
- [32] Palakshappa JA, Gibbs KW, Lannan MT, Cranford AR, Taylor SP. Systematic review of the “pragmatism” of pragmatic critical care trials. *Crit Care Explorations* 2022;4:e0738. <https://doi.org/10.1097/CCE.0000000000000738>.
- [33] Roura S, Alvarez G, Hohenschurz-Schmidt D, Solà I, Núñez-Cortés R, Bracchiglione J, et al. Lack of pragmatic attitude of self-labelled pragmatic trials on manual therapy: a methodological review. *BMC Med Res Methodol* 2024;24. <https://doi.org/10.1186/s12874-024-02393-1>.
- [34] Fritz JM, Cleland J. Effectiveness versus efficacy: more than a debate over language. *J Orthop Sports Phys Ther* 2003;33:163–5. <https://doi.org/10.2519/jospt.2003.33.4.163>.
- [35] Witt CM. Efficacy, effectiveness, pragmatic trials—guidance on terminology and the advantages of pragmatic trials. *Forsch Komple mentmed* 2009;16:292–4. <https://doi.org/10.1159/000234904>.
- [36] Gordon KS, Peduzzi P, Kerns RD. Designing trials with purpose: pragmatic clinical trials of nonpharmacological approaches for pain management. *Pain Med* 2020;21:S7–12. <https://doi.org/10.1093/pm/pnaa347>.
- [37] Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. *Cochrane handbook systematic reviews interventions*. 2nd edition. Chichester, UK: John Wiley & Sons; 2019.
- [38] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Rev Esp Cardiol* 2021;74:790–9. <https://doi.org/10.1016/j.rec.2021.07.010>.
- [39] Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. *Syst Rev* 2016;5:210. <https://doi.org/10.1186/s13643-016-0384-4>.
- [40] Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. *BMJ* 2015;350:h2147. <https://doi.org/10.1136/bmj.h2147>.
- [41] Devos F, Foissac F, Bouazza N, Ancel P-Y, Tréluyer J-M, Chappuy H. Study characteristics impacted the pragmatism of randomized controlled trial published in nursing: a meta-epidemiological study. *J Clin Epidemiol* 2019;116:18–25. <https://doi.org/10.1016/j.jclinepi.2019.07.017>.
- [42] Elder WG, Munk N. Using the pragmatic-explanatory continuum indicator summary (PRECIS) model in clinical research: application to refine a practice-based research network (PBRN) study. *The J Am Board Fam Med* 2014;27:846–54. <https://doi.org/10.3122/jabfm.2014.06.140042>.
- [43] Rosas LG, Lv N, Azar K, Xiao L, Yank V, Ma J. Applying the pragmatic-explanatory continuum indicator summary model in a primary care-based lifestyle intervention trial. *Am J Prev Med* 2015;49:S208–14. <https://doi.org/10.1016/j.amepre.2015.05.011>.
- [44] Alvarez G, Solà I, Sitjà-Rabert M, Fort-Vanmeerhaeghe A, Gich I, Fernández V, et al. A call for improving the Internal validity and the reporting of manual therapy trials self-labelled as pragmatic: a methodological review. *Int J Osteopath Med* 2025;55:100754. <https://doi.org/10.1016/j.ijosm.2025.100754>.
- [45] Bishop MD, Torres-Cueco R, Gay CW, Lluch-Girbés E, Beneciuk JM, Bialosky JE. What effect can manual therapy have on a patient's pain experience? *Pain Manag* 2015;5:455–64. <https://doi.org/10.2217/pmt.15.39>.
- [46] Lin I, Wiles L, Waller R, Goucke R, Nagree Y, Gibberd M, et al. What does best practice care for musculoskeletal pain look like? Eleven consistent recommendations from high-quality clinical practice guidelines: systematic review. *Br J Sports Med* 2020;54:79–86. <https://doi.org/10.1136/bjsports-2018-099878>.
- [47] Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ* 2014;348:g1687. <https://doi.org/10.1136/bmj.g1687>.
- [48] Dal-Ré R, de Boer A, James SK. The design can limit PRECIS-2 retrospective assessment of the clinical trial explanatory/pragmatic features. *J Clin Epidemiol* 2020;126:193–201. <https://doi.org/10.1016/j.jclinepi.2020.03.027>.
- [49] Zwarenstein M, Thorpe K, Treweek S, Loudon K. PRECIS-2 for retrospective assessment of RCTs in systematic reviews. *J Clin Epidemiol* 2020;126:202–6. <https://doi.org/10.1016/j.jclinepi.2020.06.023>.
- [50] Dal-Ré R. The PRECIS-2 tool seems not to be useful to discriminate the degree of pragmatism of medicine masked trials from that of open-label trials. *Eur J Clin Pharmacol* 2021;77:539–46. <https://doi.org/10.1007/s00228-020-03030-8>.
- [51] Dal-Ré R. Pragmatic trials, blinding, placebos, and the usefulness of the PRECIS-2 tool. *Eur J Clin Pharmacol* 2021;77:1071–2. <https://doi.org/10.1007/s00228-020-03079-5>.
- [52] Zwarenstein M, Howie A. Blinding, pragmatism, and the PRECIS-2 tool for designing and assessing randomized trials. *Eur J Clin Pharmacol* 2021;77:1069–70. <https://doi.org/10.1007/s00228-020-03078-6>.
- [53] Rossetini G, Carlino E, Testa M. Clinical relevance of contextual factors as triggers of placebo and nocebo effects in musculoskeletal pain. *BMC Musculoskelet Disord* 2018;19:27. <https://doi.org/10.1186/s12891-018-1943-8>.
- [54] Bialosky JE, Bishop MD, Penza CW. Placebo mechanisms of manual therapy: a sheep in wolf's clothing? *J Orthop Sports Phys Ther* 2017;47:301–4. <https://doi.org/10.2519/jospt.2017.0604>.
- [55] Hohenschurz-Schmidt D, Liem T. Placebo effects in osteopathy and other manual therapies – what they are and why they matter to clinical practice, education, and research. *Int J Osteopath Med* 2025;56:100762. <https://doi.org/10.1016/j.ijosm.2025.100762>.
- [56] Willis A, Shiely F, Treweek S, Taljaard M, Loudon K, Howie A, et al. Comments, suggestions and criticisms of the PRECIS-2 design tool: a citation analysis. *J Clin Epidemiol* 2024;176:111534. <https://doi.org/10.1016/j.jclinepi.2024.111534>.

5. DISCUSIÓN

5.1 Principales resultados derivados de los trabajos

Los trabajos que conforman esta tesis doctoral han permitido describir las características relacionadas con el diseño, el reporte y el riesgo de sesgo, de una cohorte de ECAs, que valoraban la efectividad de la Terapia Manual. En primer lugar, de aquellos ECAs que se autodenominaban pragmáticos en su título y/o resumen y, en segundo lugar, comparando su diseño con aquellos ECAs de efectividad pero que no se autodenominaban pragmáticos. A más a más, se ha profundizado en las características de los estudios de efectividad, diferenciando entre los ECAs que usaban un control placebo u otro tipo de grupo control.

Respecto al análisis de los ECAs autodenominados pragmáticos, obtuvimos seis resultados principales:

- 1) A pesar de la creciente atención que se presta a los estudios pragmáticos, la literatura científica no muestra un aumento de este tipo de diseño aplicado a la TM.
- 2) En segundo lugar, nuestros resultados sugieren que los estudios pragmáticos en TM tienen una actitud moderadamente pragmática. Además, la etiqueta «pragmático» podría cuestionarse en aquellos estudios que utilizan características de diseño altamente explicativas como el enmascaramiento, el entorno unicéntrico o el uso de un placebo como intervención de control.
- 3) Los autores de los estudios pragmáticos no justificaron adecuadamente por qué

el estudio pretendía ser pragmático y qué partes del diseño del estudio contribuyen más a este pragmatismo.

- 4) La calidad del reporte de los ECAs pragmáticos en TM es subóptimo.
- 5) El sesgo de “realización” y el de “detección” están presentes especialmente en los estudios pragmáticos de TM, principalmente debido a las dificultades intrínsecas de las intervenciones de TM para llevar a cabo el enmascaramiento. Además, la mitad de la muestra incurrió en un sesgo de “reporte”.
- 6) Se encontró una correlación entre la calidad del reporte y el riesgo de sesgo.

Respecto a la comparación entre los ECAs autodenominados pragmáticos con aquellos de efectividad pero que no se autodenominaron pragmáticos, obtuvimos los siguientes resultados:

- 1) El número de ECAs de efectividad han aumentado con el tiempo y presentan una actitud moderadamente pragmática. Esto sucede independientemente de si se autodenominan como pragmáticos o no y de si utilizan un grupo de control con o sin placebo.
- 2) Aunque la gran mayoría de estudios de efectividad en TM pretenden emular la práctica clínica, sus diseños no se adhieren sistemáticamente a la metodología pragmática.
- 3) El reporte de los estudios de efectividad es mayoritariamente inadecuado, y esto se agrava cuando los estudios no se autodenominan pragmáticos. También es más deficiente cuando se utiliza el “usual care” u otras intervenciones como intervención control en lugar de utilizar un control placebo.

- 4) El sesgo de realización está presente en prácticamente toda la muestra. Si bien el sesgo de selección es menor cuando los estudios se autodenominan pragmáticos, el sesgo de detección es menor cuando se usa un control placebo que cuando se usa otra intervención activa o “usual care”.
- 5) El uso de la terminología eficacia/efectividad parece no estar claro entre la comunidad investigadora en TM, usando el término efectividad en diseños con características explicativas.

5.2 Discusión de los elementos más relevantes de los estudios de efectividad

5.2.1 Estado de la investigación sobre la efectividad de la Terapia Manual

Aunque la investigación en el campo de la TM ha aumentado en las últimas décadas [38,59,131], la revisión de la literatura muestra que los estudios pragmáticos no son tan comunes. Sólo 39 artículos que evaluaban una intervención de TM se identificaron como pragmáticos hasta 2024. Nuestros resultados indican que el número de EPs en TM no ha aumentado en los últimos años, y la puntuación media obtenida con la herramienta PRECIS-2 tampoco ha aumentado con el tiempo. Estos resultados son contrarios a los estudios que evalúan el pragmatismo en los cuidados intensivos [108], el dolor [32] y las intervenciones cardiovasculares [102], en los que la puntuación media del PRECIS-2 aumentó con el tiempo. Las posibles causas para entender esta falta de aumento de estudios pragmáticos a lo largo del tiempo en el campo de la TM o de las terapias físicas podría ser por la falta de conocimiento que tienen los investigadores en este campo sobre los estudios pragmáticos para intervenciones no farmacológicas y/o la falta de formación específica sobre este tipo de estudios en las etapas académicas.

No obstante, los estudios de efectividad que pretenden asemejarse a la práctica clínica pero que no se autodenominan «pragmáticos» sí que han aumentado en los últimos

años. El hecho de que en la comunidad científica hubiera múltiples llamamientos para mejorar la aplicabilidad y el diseño de los estudios para evaluar el efecto de las intervenciones en un contexto clínico real puede haber impulsado el número de estudios de «efectividad» [43,85,94,96].

5.2.2 Actitud pragmática y características metodológicas de los ECAs de efectividad en Terapia Manual

Valorar la actitud de un estudio es fundamental porque establece la base para comprender sus objetivos, evaluar la adecuación de su diseño y determinar la relevancia y aplicabilidad de sus resultados para la toma de decisiones en la práctica clínica real. No se trata solo de clasificar los estudios, sino de proporcionar el contexto necesario para interpretar la evidencia de manera informada.

La puntuación media en la evaluación retrospectiva con la herramienta PRECIS-2 de los estudios autodenominados pragmáticos en nuestra muestra, fue de 3.5 (DE=0.6), lo que indica un nivel moderado de pragmatismo. Puntuaciones similares se encontraron en una revisión sistemática de intervenciones de enfermería [103] y en intervenciones sobre el dolor [32] y otros campos [98,102,108,109,132]. El nivel moderado de pragmatismo encontrado en nuestra revisión estuvo muy influenciado por los dominios «flexibilidad de la intervención», «seguimiento», «variable de resultado» y «análisis primario». Estos dominios se calificaron como pragmáticos (4 o 5 en la escala de Likert) en más del 60% de los estudios evaluados. El uso de un análisis por intención de tratar es un rasgo altamente pragmático, y la mayoría de las revisiones que evalúan el pragmatismo de los estudios en otros campos encontraron este ítem como el más pragmático en su evaluación con la herramienta PRECIS-2 [32,102,103,108]. Por el contrario, los estudios de efectividad no autodenominados pragmáticos obtuvieron una puntuación media PRECIS-2 de 2.8, significativamente inferior a la de los estudios de efectividad

autodenominados como pragmáticos (3,5). Especialmente el dominio 3 (entorno donde se realiza el estudio) y el 5 (intervención) fueron explicativos en casi toda la muestra. Este hecho podría revelar que cuando se elige un diseño pragmático y se autodenomina como tal, se presta más atención al diseño del estudio con características pragmáticas.

El uso de controles con placebo, los contextos unicéntricos y el enmascaramiento, que son características muy explicativas, ha dado lugar a un debate sobre si los estudios con estas características deberían considerarse directamente explicativos y no valorar su pragmatismo [46-50]. En este caso, deberíamos haber excluido casi la mitad de nuestra muestra (19/39). Sin embargo, como no existe un consenso al respecto, decidimos evaluar toda la muestra y seguir contribuyendo al debate como hicieron otros autores en otros campos [32,103,109,132].

5.2.2.1 Elementos explicativos en estudios pragmáticos de Terapia Manual

Ciertas características de diseño (como el uso de placebo, el cegamiento, o el ámbito unicéntrico del estudio) pueden ser más o menos apropiadas dependiendo de la actitud del ensayo. Valorar la actitud del estudio de una forma no dicotómica proporciona el contexto necesario para juzgar la idoneidad de estas características y evitar categorizar prematuramente un ensayo basándose en una sola característica del diseño. A continuación, describimos las particularidades de estas características consideradas explicativas en el contexto de la terapia manual.

Placebo

El uso de un placebo como comparador se considera una característica muy explicativa. Algunos autores argumentan que el placebo en ningún caso forma parte de las intervenciones utilizadas en la práctica real, y que, por tanto, su uso sugiere una naturaleza explicativa en el diseño [111,117,118,133]. En cambio, los autores de PRECIS-2 defienden que hay situaciones en las que un estudio podría incluir un control placebo y aun así proporcionar información útil para la toma de decisiones (por ejemplo,

para mitigar la subjetividad de los resultados informados por los pacientes en un estudio de efectividad comparativa de dos tratamientos aceptados) [46,116]. Diversos estudios han destacado el papel del efecto placebo de las intervenciones manuales como parte activa de los efectos del tratamiento [12,13,29,30]. La inclusión de un grupo control placebo puede tener connotaciones explicativas. No obstante, su inclusión en un ECA de efectividad de la TM podría tener implicaciones importantes. Principalmente, para comprender el efecto real de las intervenciones, difícil de conseguir en un estudio de eficacia realizado en condiciones no reales (donde no están presentes todos los factores contextuales que forman parte del encuentro entre el paciente y el terapeuta en este campo). Del mismo modo, incluir un grupo placebo ayudaría a entender si el efecto de la intervención es debido a la propia intervención o a otros factores como las expectativas del paciente, la regresión a la media o la evolución natural de la enfermedad [54]. Entre los estudios pragmáticos analizados, sólo un estudio utilizó un control con placebo [134], y otro incluyó un control con placebo en un estudio de 3 brazos [135]. En este último, Mafetoni et al. mostraron que, aunque el efecto de la intervención de TM fue mayor de forma significativa que en el grupo control con usual care y que el grupo con control placebo, el grupo placebo mostró un efecto mayor que el grupo control [135]. De este modo, el estudio no sólo corroboró la existencia del efecto placebo y su efecto terapéutico, sino que consiguió aislar el verdadero mecanismo de acción de la intervención activa, explicando por qué su eficacia es superior. Es importante matizar que estos dos estudios puntuaron 3,8 y 3,6 respectivamente en el sumatorio de los dominios de PRECIS-2, ligeramente por encima de la media global de todos los estudios analizados, lo que podría justificar la idoneidad de poder incluir un grupo control placebo en los estudios pragmáticos.

Las intervenciones no farmacológicas no siguen los mismos procesos de desarrollo e implementación que las intervenciones farmacológicas [82]. Este hecho conlleva que, a menudo, o bien se realicen estudios de efectividad para evaluar las intervenciones sin haber realizado antes estudios de eficacia, o bien, que se hayan realizado estudios de eficacia no concluyentes donde no se han podido aislar los mecanismos de acción de

estas intervenciones, pero que son intervenciones con claros resultados clínicos. Algunos autores plantean que, en estas situaciones, con intervenciones ampliamente establecidas en la práctica clínica, realizar estudios pragmáticos podría ser una solución a la falta de evidencia, aunque sus resultados deben interpretarse con cautela [21,98]. En estos casos, incluir un grupo control placebo podría ser muy adecuado para objetivar la magnitud real del efecto de la intervención. Nuestra revisión reveló que no hubo variaciones significativas en la evaluación con la herramienta PRECIS-2 entre los estudios que emplearon un placebo y los que emplearon otras intervenciones o el “usual care”, obteniendo el mismo grado de pragmatismo “moderado” independientemente del grupo control utilizado. Estos hallazgos coinciden con los encontrados por Devos et al. analizando el pragmatismo de intervenciones de enfermería [103] y contradicen la lógica planteada por Dal-Ré et al. de que el placebo se aleja de la práctica real y debería reducir el pragmatismo del estudio [117]. Si bien se está considerando la posibilidad de incluir un nuevo ítem en el desarrollo de la futura herramienta PRECIS-3 donde se evalúe el grupo control [99], deberían realizarse más investigaciones para contribuir a este interesante debate sobre el uso de placebo en los estudios pragmáticos [117–119].

Ámbito de estudio unicéntrico

Otro punto controvertido en los estudios pragmáticos es el hecho de realizar la intervención en un único centro, ya que existe un debate sobre si estos estudios deberían considerarse pragmáticos por la discutible generabilidad de sus resultados a otros entornos [111,116,118,119]. Nuestros resultados mostraron que un alto porcentaje de estudios de efectividad fueron realizados en un solo centro. La realización de un estudio en un único centro afecta, sin duda, a la generabilidad de los resultados del estudio [111]. No obstante, como destacan Zwarestein et al., si en el estudio se especifican detalles relativos al entorno, como el tipo de sistema sanitario, las normas regulatorias sanitarias y los tipos de pacientes, un entorno unicéntrico puede seguir proporcionando información relevante para la práctica clínica [116]. Cabe destacar que los estudios analizados no contextualizaron estos aspectos relativos al entorno que facilitan la interpretación de su generabilidad. Además, la realización de estudios multicéntricos

acarrea gastos considerables y dificultades logísticas, lo que supone un reto para los investigadores de TM, campo en el que no hay intereses comerciales que favorezcan la financiación de estos estudios.

Enmascaramiento

La consideración del enmascaramiento en los estudios pragmáticos también es un tema de debate [117–119]. Algunos autores plantean la limitada capacidad que tiene la herramienta PRECIS-2 para distinguir entre estudios enmascarados y abiertos, poniendo en duda su utilidad en estos casos [117]. Al igual que el uso de controles placebo, el enmascaramiento se aleja de la práctica clínica habitual, por lo que debería reducirse el grado de pragmatismo de los estudios que enmascaran a los pacientes y/o terapeutas. En el caso de la TM, el enmascaramiento de los terapeutas no es posible y el de los pacientes es particularmente complicado. Del mismo modo, cuando se usan variables de resultado reportadas por los pacientes, el enmascaramiento del evaluador externo es discutible puesto que son los propios pacientes quienes reportan los resultados. Esto dibuja una situación muy proclive al pragmatismo en el campo de la TM, pero que a la vez implica un riesgo de sesgo elevado. La evaluación de estos estudios mediante el uso de herramientas de riesgo de sesgo (como la herramienta Cochrane RoB), penaliza la validez interna del estudio disminuyendo considerablemente su calidad o certeza de la evidencia. Si bien el enmascaramiento puede ser deseable, un diseño pragmático que considere cuidadosamente la selección de pacientes, las intervenciones y las variables de resultado, puede ayudar a reducir el sesgo incluso en ausencia de enmascaramiento [136]. Por ejemplo, la combinación de medidas objetivas de resultado que puedan relacionarse con las medidas reportadas por los pacientes sería recomendable en estos casos.

Todos estos aspectos relacionados con elementos explicativos dentro de los estudios pragmáticos, se están considerando para el desarrollo de la nueva herramienta PRECIS-3 donde, además de incluir nuevos ítems y clarificar los ya existentes, se valorará el otorgar más peso a unos dominios u otros para poder cuantificar de manera más fiable

el pragmatismo de un estudio [99].

5.2.2.2 Intervención experimental

En los estudios de efectividad (tanto si se autodenominan pragmáticos como no), el objetivo principal es evaluar la efectividad comparativa de una intervención en entornos de la práctica real, donde la flexibilidad es crucial para su aplicabilidad y generalización a diferentes contextos. Si bien protocolizar rígidamente la intervención experimental favorece la reproducibilidad de los estudios, puede comprometer varios aspectos clave de la atención clínico-terapéutica [41]. En la práctica clínica, así como en las recomendaciones de la literatura, las Terapias Manuales actúan bajo un paradigma centrado en el paciente [4,137,138]. El paradigma de la atención centrada en el paciente (patient-centered care) debe ser considerado en la investigación no farmacológica, donde las intervenciones no siempre se implementan de forma uniforme. Adaptarse a las diferencias individuales de los pacientes, a sus preferencias, a las características de su entorno y/o a los recursos disponibles, puede aportar resultados más relevantes y aplicables a la práctica clínica [48]. Además, si la intervención se protocoliza de forma estricta o rígida, es menos probable que los resultados del estudio sean aplicables a otros entornos y poblaciones y/o que reproduzcan de forma fiel lo que sucede en la práctica clínica habitual [18,55,56]. Negrini et al. pusieron en evidencia la brecha existente entre la investigación y la clínica en cuanto a la replicabilidad clínica de las intervenciones complejas, encontrando que sólo el 16% de los ECAs sobre intervenciones de rehabilitación incluidos en su estudio fueron considerados aplicables por los clínicos [139].

En nuestra revisión, la intervención utilizada con mayor frecuencia en los estudios pragmáticos (más del 50%) fue una combinación de técnicas no protocolizada. En cambio, en los estudios de efectividad, se utilizó con mayor frecuencia una combinación de técnicas protocolizada (en los estudios que usaban una intervención control activa) y una técnica aislada protocolizada (en los estudios con control placebo). Estos resultados

concuerdan con los encontrados por Cottone et al. donde evaluaron el abordaje pragmático de las técnicas manipulativas y de movilización en pacientes con dolor cervical [140]. En este estudio, el dominio “flexibilidad de la intervención” puntuó valores que tendían a la eficacia más que a la efectividad. Una técnica aislada o un protocolo de técnicas respondería a un estudio de eficacia donde el objetivo podría ser comprobar el efecto fisiológico que provoca una técnica en un tejido o sistema en concreto, a menudo, a expensas de alejarse de la realidad clínica. No obstante, protocolizar las intervenciones puede ser muy interesante en algunas ocasiones, como, por ejemplo, para comprobar dentro de los diferentes tipos de técnicas de TM, qué técnica ofrece mejores resultados en diferentes grupos de pacientes [48]. Pero en este caso, de nuevo, la intención del estudio sería más explicativa que pragmática.

La decisión de no estandarizar las intervenciones supone un reto en investigación ya que dificulta la replicación de los estudios, premisa imprescindible para realizar un ECA. Sin embargo, la flexibilidad en la intervención no implica la ausencia de un protocolo o de una descripción detallada de la misma [18]. El uso de guías de reporte específicas para la descripción de las intervenciones como TIDieR [141] son imprescindibles y su adherencia debería estar asegurada por parte de los editores y revisores por parte de las revistas científicas en el proceso de publicación de los estudios. Siendo esta recomendación válida para cualquier tipo de estudio, es especialmente importante en estudios de efectividad sobre intervenciones no farmacológicas. Si bien la adaptación de la intervención y el manejo del paciente es razonable para asegurar la validez externa del estudio en el contexto de la atención centrada en el paciente, dichas adaptaciones deben reportarse con el suficiente detalle para asegurar al máximo su replicabilidad. La clave está en encontrar un equilibrio entre la estandarización necesaria para asegurar la fidelidad a la intervención y la flexibilidad requerida para adaptarla a las condiciones de la práctica clínica real [18,138].

Para ello, diversos autores han propuesto distintas estrategias:

- 1) Definir los componentes esenciales de la intervención: aunque no se protocolice la intervención, se pueden describir aquellos componentes que se tendrán en cuenta en todos los contextos para asegurar su efectividad [142].
- 2) Establecer límites claros para las adaptaciones que se realizarán [21,143].
- 3) Documentar y monitorear rigurosamente las adaptaciones que se realizan en cada contexto [16,32,56,142].

5.2.2.3 Intención, contexto y justificación de los estudios de efectividad

Recientemente, se ha señalado que, al evaluar la actitud pragmática en un estudio, hay que prestar especial atención a detectar su intención, además de otorgar una puntuación cuantitativa de su pragmatismo [50,111]. Existe una diferencia entre la «actitud» de un estudio y su «intención», aunque ambas estén estrechamente relacionadas. Según Schwartz y Lellouch [1], la «actitud» de un estudio se refiere al enfoque general que guía el diseño y la realización de un estudio clínico. Existen dos actitudes principales: 1) Actitud explicativa: Centrada en la comprensión de los mecanismos y la comprobación de hipótesis biológicas. Y 2) Actitud pragmática: Orientada a la toma de decisiones, con el objetivo de determinar qué tratamiento es preferible en la práctica.

En cambio, la «intención» de un estudio es el objetivo específico o la pregunta que pretende responder el estudio. La **actitud** de un estudio responde al “por qué” general de la investigación: si el objetivo es comprender un mecanismo (actitud explicativa) o tomar una decisión práctica (actitud pragmática). La **intención**, en cambio, define con nitidez el “qué” se va a investigar. Por ejemplo, en un estudio de terapia manual, una **Actitud explicativa**: investiga el mecanismo de acción; y la **Intención**: determinar si una técnica concreta de TM se correlaciona con cambios en la variabilidad de la frecuencia cardíaca. O en otro ejemplo, una **Actitud pragmática**: evalúa la utilidad clínica; y la **Intención**: comparar si, en condiciones reales de práctica, añadir TM al tratamiento estándar de la infección pulmonar obstructiva mejora los resultados respecto al

tratamiento estándar solo.

Así, pues, aunque la actitud elegida da forma e informa la intención, no son idénticas. La actitud marca la finalidad general del estudio y orienta el enfoque, mientras que la intención traduce esa orientación en un objetivo de investigación concreto. Es importante diferenciar ambos términos y evaluar en los estudios cómo el autor ha descrito la intención del estudio más allá de calificar el diseño del estudio de pragmático o explicativo [50].

Menos del 20% de los estudios autodenominados pragmáticos en nuestra revisión informaron que el objetivo era investigar una intervención en condiciones de la práctica real y tomar decisiones clínicas sobre su efectividad, por lo que no se identificó una intención completamente alineada con lo que debería ser una actitud pragmática. Por lo que, casi el 80% de los estudios auto-denominados pragmáticos no justificaron su intención pragmática. Tampoco dieron explicaciones sobre qué dominios se diseñaron con el fin de garantizar la validez externa y cuáles para garantizar la validez interna del estudio.

Estos resultados observados en nuestra revisión, coinciden con el estudio de Janiaud et al. donde evaluaron la actitud pragmática de los estudios pragmáticos en diferentes campos, y concluyeron que el 45% de los autores no justificaban ni discutían adecuadamente el pragmatismo alegado [132].

Al conocer la intención del estudio, cabe esperar que algunos dominios puedan ser más pragmáticos que otros. Algunos autores afirman que la actitud pragmática es específica de cada dominio y que las puntuaciones de la herramienta PRECIS-2 requieren una justificación para entender cada dominio en el contexto del estudio [32]. De hecho, uno de los principios fundamentales de la validez externa es que el contexto importa [43,112], ya que puede ayudar a comprender mejor la actitud del estudio, evitando la dicotomización en pragmático y explicativo [50]. Los autores de la herramienta PRECIS-

2 no recomendaban cuantificar el pragmatismo de los estudios, sino que apoyaban el continuo explicativo-pragmático valorando cada dominio en relación a las necesidades del estudio [95]. Los comentarios recibidos por diversos autores que han usado la herramienta de forma retrospectiva [32,103,104,109,117,118] ha hecho que, de cara a la futura herramienta PRECIS-3 [99], se plantee la idea de darle una ponderación específica a cada dominio que otorgue un peso distinto al pragmatismo total del estudio. Por nuestra parte, consideramos que más allá de esta cuantificación para los dominios PRECIS-3, los autores de los estudios deberían, además, justificar claramente qué áreas del estudio reflejan la práctica habitual. A la vez, justificar aquellos elementos que, por requisitos del estudio, comprometen su validez externa y en qué grado. Estas especificaciones facilitarían la comprensión del estudio por parte del lector y ayudaría a entender su grado de aplicabilidad.

La intención está directamente relacionada con la pregunta de investigación y debe ser expuesta por los autores. ¿El estudio pretende demostrar la eficacia de una intervención en condiciones ideales o en un entorno similar al clínico? ¿El estudio pretende parecerse a la práctica clínica o los autores pretenden probar los efectos fisiológicos de una intervención independientemente de si el entorno del estudio es apropiado o no? Conociendo la intención, el diseño y los datos que proporcionen los resultados del estudio y su aplicabilidad diferirán [45]. En nuestra revisión, identificamos discrepancias entre los objetivos de los estudios y su metodología [144,145]. Esto coincide también con las observaciones realizadas por Maddox et al. en su reciente revisión sistemática [48] donde observaron que la mayoría de estudios que investigaban el efecto de la TM en el dolor lumbar carecían de pragmatismo. Y también con los resultados de Cottone et al. donde concluye en su revisión sistemática que los estudios que estudian la efectividad de las movilizaciones y manipulaciones articulares para el dolor cervical suelen tener diseños explicativos que carecen de aplicabilidad [140].

Tanto los estudios de eficacia como los de efectividad son necesarios, pero abordan cuestiones de investigación diferentes [45,48]. En el momento de plantearse la realización

de un estudio pragmático, una pregunta potencial a hacer es si ya existen suficientes estudios de eficacia para explorar más a fondo los efectos de una intervención en un entorno más real. Idealmente, los autores deben examinar la bibliografía existente para evaluar lo que ya se sabe sobre una determinada intervención antes de plantear un estudio pragmático. Es decir, debería proporcionarse una base teórica sólida junto con la justificación para realizar un estudio pragmático [45,142]. Dicha justificación podría incluir experimentos mecanicistas existentes y estudios de eficacia controlados con placebo, que sugieran que la intervención tiene los efectos fisiológicos y la eficacia deseados [126]. Sin embargo, en el 22% de los estudios pragmáticos y el 25% de los estudios de efectividad (no autodenominados pragmáticos) que revisamos, los autores no describieron ninguna teoría sólida de la intervención ni el estado de los estudios previos que abordaban cuestiones de investigación relacionadas (justificación de la intervención). Además, cuando se proporcionó la justificación, esta se construyó principalmente a partir de estudios de efectividad comparativa. Estos resultados coinciden con los que encontraron Taljaard et al. donde revisaron sistemáticamente estudios que tuvieran la etiqueta “pragmático” en referencia al diseño del estudio y valoraron como los autores usaban esta etiqueta [109]. En ella se observó que el 71,3% de la muestra no justificaba por qué se había seleccionado un estudio pragmático en lugar de uno explicativo [109]. Consideramos que estas reflexiones son fundamentales cuando se pretende idear y diseñar un estudio pragmático [21].

Es fundamental que los autores de un estudio de investigación describan explícitamente la intención, el contexto y la justificación de sus estudios, más allá de simplemente categorizarlos como pragmáticos o explicativos, para de esta forma ayudar a los lectores a interpretar y aplicar con mayor criterio los resultados del estudio.

5.2.3 Reporte de los estudios de efectividad

Globalmente, el reporte de los estudios de efectividad es considerado deficiente, a pesar de la existencia de guías de reporte, y siendo este esencial para la aplicabilidad y la

calidad de los resultados [45,65].

Aproximadamente el 30% de los estudios incluidos en nuestra revisión declararon seguir alguna guía de reporte. Además, el grado de cumplimiento de los ítems de la guía CONSORT fue ligeramente mayor en los estudios pragmáticos que en los estudios de efectividad no autodenominados pragmáticos. Sólo un estudio [130] reportó haber usado la extensión CONSORT para estudios pragmáticos [97], que sugiere que su uso todavía no es habitual a pesar de las recomendaciones en este campo [106].

La calidad del reporte de nuestra muestra fue, en general, subóptimo. Los estudios de efectividad con una intervención no placebo, se informaron peor que los estudios pragmáticos y que los estudios de efectividad controlados con placebo. Sólo el 25% de nuestra muestra reportó el 70% de los ítems recomendados por la guía CONSORT, y el 30% reportó de forma aceptable menos del 50% de los ítems. Estos resultados coinciden con los de otros autores que evaluaron el reporte de ECAs de eficacia en TM [1] y rehabilitación [65]. Sin embargo, analizando en profundidad esta comparación, observamos que los ítems de reporte relacionados con las secciones de métodos y resultados se reportaron ligeramente mejor en nuestra muestra focalizada en estudios pragmáticos. Por ejemplo, la descripción del tamaño de la muestra, la asignación aleatoria y secuencial, y la descripción de los resultados (diagrama de flujo). Los motivos de esta ligera mejoría en el reporte de los estudios pragmáticos en comparación con los ECAs de efectividad no auto-reportados como pragmáticos puede ser debido, en primer lugar, a la existencia de la guía “CONSORT Extension for Pragmatic Trials” [97], que obliga a describir con precisión el contexto real, el flujo de participantes y las adaptaciones de la intervención en este tipo de diseño. Además, su énfasis en la validez externa y la aplicabilidad clínica puede ser el motivo por el cual los autores intenten documentar de forma más precisa detalles metodológicos del estudio. Por último, la flexibilidad inherente al diseño pragmático, en contraste con los ensayos explicativos más rígidos, puede ser que genere una variabilidad que deba registrarse y explicarse con detalle.

Un elemento fundamental para comprender el pragmatismo de un estudio es el análisis de cómo se ha realizado la intervención y cuáles son sus características. Esto es importante para todos los estudios de efectividad, se hayan o no auto-denominado pragmáticos y es especialmente relevante en una intervención compleja no farmacológica como es la TM [131]. En nuestra muestra, el reporte del ítem 5 de la guía CONSORT (Intervenciones) fue, en general, deficiente. Esta es una observación ya reportada en la literatura [1]. Aunque se proporcionaron descripciones generales de las intervenciones, detalles como el procedimiento utilizado para adaptar la intervención a los participantes de forma individual y la forma como se evaluó la adherencia de los participantes y proveedores estuvieron presentes en menos de la mitad de nuestra muestra. Estos puntos son especialmente relevantes para que los lectores no familiarizados con la intervención comprendan el contexto del estudio y poder así garantizar la reproducibilidad de la intervención [146]. Si bien una intervención protocolizada disminuye los sesgos y facilita la replicación del estudio [58], en el caso de la TM dichos protocolos se contraponen al uso habitual que se hace de ella clínicamente y al enfoque centrado en la persona. En la aplicación de la TM, el tratamiento puede diferir entre pacientes debido a la heterogeneidad en la forma de presentación y la fase de las patologías. Por ejemplo, el dolor lumbar inespecífico puede tener una etiología múltiple en función de cada paciente, por lo que requiere de enfoques individuales y de diferentes técnicas manuales según la evolución de la afección [4,74]. En estos casos, comunicar la información sobre cómo se adaptará la intervención a cada participante facilitará la replicación del estudio y aumentará directamente la calidad metodológica del mismo [18,58]. De hecho, este punto (adaptación de la intervención) se incluye en la guía TIDIER [141] la cual ha sido especialmente recomendada para el campo de la TM [131]. Sin embargo, ningún estudio de nuestra muestra mencionó el uso de la lista de verificación TIDIER u otras herramientas de reporte similares para describir la intervención. A pesar de varios llamamientos para mejorar el reporte de la intervención [141,147,148], una reciente overview de revisiones sobre el reporte de la investigación sanitaria concluyó que este era inadecuado [149]. El reporte exhaustivo de cualquier

intervención de investigación en estudios clínicos es esencial para evaluar la relevancia de los hallazgos en el contexto de la práctica habitual [131,150].

El análisis de nuestra muestra detectó un reporte adecuado sobre la forma en que se reportó el enmascaramiento y las estrategias empleadas para limitar el sesgo de detección cuando éste no fue posible. No obstante, aquellos estudios que utilizaron un control placebo mostraron un mejor reporte en todos los ítems relacionados con el enmascaramiento que los que usaron otro tipo de intervención control. Reportar cómo fue el enmascaramiento y qué estrategias se emplearon en caso de que el enmascaramiento no fuera posible está altamente recomendado en la extensión CONSORT para estudios no farmacológicos [82,127] y la extensión CONSORT para estudios pragmáticos [97] para mitigar el posible sesgo producido por la falta de enmascaramiento.

Cabe destacar que los ítems acerca de los «daños» y el «número de registro del ensayo y dónde puede evaluarse el protocolo» no se informaron en casi el 50% de la muestra. No registrar el protocolo de los ensayos conlleva un potencial sesgo de reporte que afectará a la evaluación de la calidad de los estudios.

5.2.4 Riesgo de sesgo de los estudios de efectividad

Además de evaluar el pragmatismo de los estudios de Terapia Manual, uno de nuestros objetivos secundarios fue analizar el riesgo de sesgo que prestaban estos estudios. Evaluar el riesgo de sesgo en los estudios de efectividad de una intervención de terapia manual resulta crucial para garantizar que las conclusiones extraídas sean fiables y aplicables a la práctica clínica. Un sesgo no identificado o mal controlado, como la falta de aleatorización, la ausencia de cegamiento de los evaluadores o las pérdidas diferenciales de participantes entre los grupos de estudio, puede distorsionar los resultados y generar expectativas injustificadas sobre el beneficio real de la terapia. Además, un análisis riguroso del sesgo permite establecer la calidad del estudio y poder

así comparar de manera más precisa y transparente diferentes investigaciones, favoreciendo la toma de decisiones.

Cabe destacar que el sesgo de selección fue ligeramente menor en los estudios autodenominados pragmáticos. Resultados similares encontraron Gastaldon et al. en una revisión sistemática donde evaluaban las características metodológicas de los estudios pragmáticos en intervenciones psicológicas y psicosociales en pacientes con esquizofrenia y psicosis [151]. Este estudio mostró que los estudios pragmáticos presentan, en general, un riesgo de sesgo igual o incluso menor que los estudios explicativos, lo que desafía la idea de que validez interna y externa sean mutuamente excluyentes [151].

No obstante, nuestra evaluación con la herramienta RoB encontró sesgos de realización y detección en casi toda la muestra, tanto en los estudios autodenominados pragmáticos como en los que no se autodenominaban pragmáticos, y tanto los que utilizaban un control placebo como los que utilizaban otra intervención como control. Debido a las dificultades inherentes para implementar estrategias de enmascaramiento efectivas en los estudios de TM, el sesgo de realización supone un verdadero desafío en la investigación de las intervenciones de TM [4,18,111]. De forma general, estudios meta-epidemiológicos, sugieren que existe una tendencia hacia la sobreestimación de los efectos del tratamiento en los ECAs para ciertos tipos de sesgo, especialmente cuando la calidad metodológica es inadecuada, y especialmente en el campo de la Rehabilitación [152,153] No obstante, las conclusiones de estos estudios son poco concluyentes y los autores concluyen que se requiere de más investigación en este campo para poder determinar la influencia exacta de la existencia de sesgos en la estimación de los efectos de la intervención [152,153].

Si bien, en otros campos de la medicina, parece que el hecho de no cegar a los pacientes probablemente sobreestima los efectos del tratamiento para las variables PROMs [153], otros autores defienden que el enmascaramiento podría ser menos importante de lo que

se creía para estimar los efectos de las intervenciones [154,155].

De hecho, la necesidad de enmascaramiento en los estudios de efectividad es un tema de debate, como ya se ha mencionado anteriormente [117–119]. En su lugar, se han propuesto estudios abiertos y recomendaciones como las de enmascarar a los evaluadores externos para que los estudios se asemejen más a la práctica clínica [156]. Para la evaluación de intervenciones centradas en el paciente, una opción altamente recomendada es el uso de variables auto-reportadas por los propios pacientes (PROMs) [157]. En estos casos, el paciente se convierte en el evaluador externo. Si no ha sido posible cegar adecuadamente a los pacientes, resulta imposible enmascarar a los evaluadores de los resultados, lo que conduce de nuevo al sesgo de detección. La declaración IMMPACT publicada recientemente sobre objetivos y consideraciones de la investigación en estudios pragmáticos de intervenciones para el dolor, propone diferentes estrategias para limitar estos sesgos, como la realización de estudios de efectividad comparada sin enmascaramiento, el enmascaramiento de la hipótesis del estudio o los diseños Zelen [142].

Por otro lado, existe un llamamiento urgente para que los investigadores en TM registren prospectivamente los protocolos de investigación y reporten si el protocolo existe y dónde se puede encontrar. El sesgo de reporte no estaba claro en la mitad de nuestra muestra por este motivo.

Todos estos sesgos están descritos en la literatura sobre TM [4,111,156] y afectan en gran medida a la calidad percibida de cualquier ECA y a la credibilidad de sus resultados. Ante esta situación, algunos autores podrían cuestionar la sensibilidad de la herramienta RoB para evaluar la calidad de un estudio que evalúa una intervención compleja no farmacológica como la TM, especialmente si el estudio ha sido diseñado con una actitud claramente pragmática y sigue sus premisas. Sin embargo, pensamos que, más allá de la sensibilidad de la herramienta, es la propia naturaleza de estas intervenciones y la calidad de su reporte lo que convierte en un gran desafío la aplicación e interpretación

de los resultados de la herramienta RoB para valorar la calidad de los estudios de TM.

5.4.5 Relación pragmatismo-reporte-riesgo de sesgo

Relacionar el grado de pragmatismo de un estudio con la calidad de su reporte y su riesgo de sesgo resulta esencial para valorar hasta qué punto las intervenciones son aplicables al entorno clínico real. Además, analizando esta información se puede investigar la relación entre validez interna y externa. Tradicionalmente, se ha sugerido que los ensayos explicativos priorizan la validez interna (calidad/bajo riesgo de sesgo) a expensas de la validez externa (pragmatismo/aplicabilidad al mundo real) [151]. Es importante identificar qué dominios específicos están peor reportados y averiguar cómo este aspecto afecta la capacidad de juzgar tanto el pragmatismo como la calidad metodológica. Evaluar conjuntamente estos tres aspectos permite equilibrar la validez interna (control de sesgos) con la validez externa (generabilidad), asegurando que los resultados reflejen tanto la rigurosidad metodológica como la factibilidad de su implementación en entornos de práctica real.

Nuestro análisis no mostró una relación significativa entre el grado de pragmatismo, la calidad del reporte y el riesgo de sesgo. No obstante, sí concluimos que un buen reporte de los ítems CONSORT estuvo directamente relacionado con un “bajo RoB” en todos los estudios de efectividad en TM. De hecho, en TM y rehabilitación física, estudios recientes de meta-investigación han mostrado una alta correlación entre el buen reporte y la calidad de los estudios [1,65,151]. Un mal reporte dificulta tanto la valoración de los dominios clave de pragmatismo (como la elegibilidad de los participantes, la flexibilidad de la intervención o las variables de resultado) como la cuantificación de los sesgos existentes (selección, realización, detección, abandono). Mejorar la calidad del reporte de los ECAs es básico para poder valorar de forma transparente el pragmatismo de los estudios y analizar rigurosamente la presencia o no de sesgos en los estudios y ver si estos influyen en la estimación del efecto de las intervenciones y, consecuentemente, en la fiabilidad de los resultados reportados [151–153].

5.3 Idoneidad en el uso de la terminología eficacia/efectividad en los estudios de TM

Nuestra revisión identificó una tendencia creciente en los estudios de efectividad diseñados para reflejar la práctica clínica. Sin embargo, al evaluar el pragmatismo de distintas cohortes de estudios se constata que, en el campo de la TM y la rehabilitación, los estudios de efectividad carecen de la validez externa necesaria para garantizar su aplicabilidad en entornos más reales [48,140,145]. La combinación del aumento de los estudios de efectividad y la limitada aplicabilidad de sus resultados, evidencia que la distinción entre estudios de eficacia y efectividad aún resulta poco clara [45,49,51,52,158]. Dos factores clave definen la posición de un estudio dentro del continuo eficacia-efectividad: la intención del estudio [50] y la justificación de su diseño [45]. La intención determina la pregunta de investigación, afecta al diseño e influye en los resultados y la aplicabilidad del estudio [45,50]. La justificación debe determinar si hay suficientes ensayos de eficacia que justifiquen la evaluación en entornos más reales [45]. Nuestro análisis reveló que el 25% de los ensayos carecían de esta justificación y los que la tenían a menudo se basaban en experimentos mecanicistas en lugar de ensayos controlados con placebo o tratamiento simulado (sham). Este desajuste entre los objetivos planteados y la metodología empleada, evidenciado en el uso impreciso de la terminología eficacia/efectividad, ha sido señalado en estudios previos [45,158,159].

En cuanto al uso de una herramienta de diseño para orientar el diseño de un estudio en el continuo explicativo-pragmático, sólo un estudio de nuestra muestra informó del uso de la herramienta PRECIS-2 [160]. Cabe destacar que este estudio obtuvo una puntuación media de pragmatismo de 4.1 sobre 5, significativamente superior a la de los demás estudios analizados. El uso limitado de este tipo de herramientas de diseño ha

120

sido también evidenciado en investigaciones recientes [102,103,108], lo que subraya la necesidad de integrar estas herramientas de forma sistemática en la planificación y evaluación de estudios clínicos.

Esta incertidumbre conceptual y terminológica detectada tiene implicaciones importantes. No solo complica la interpretación de los resultados de los estudios clínicos, sino que también puede afectar a las recomendaciones incluidas en las guías de práctica clínica [51,52]. Tal y como Porzsolt et al. constatan en su revisión: *“El uso erróneo de la terminología no es sólo un problema académico; puede tener importantes implicaciones para la comunicación médica. El uso incoherente de la terminología socava tanto la investigación básica como la aplicada. La confusión entre investigación sobre eficacia y efectividad puede dificultar la interpretación de los estudios clínicos y la comunicación médica de los estudios clínicos y la información que contienen las guías de práctica clínica”* [51].

Dal-Ré et al. reflexionan sobre cómo los estudios de intervenciones no reguladas (que no pasan procesos de regulación como los medicamentos) puede haber contribuido al uso impreciso de la terminología eficacia/efectividad ya que en estos estudios la distinción entre estos conceptos resulta difusa. Esto dificulta poder determinar si se está evaluando la eficacia o la efectividad en intervenciones tan diversas como nuevas técnicas quirúrgicas, fisioterapia o intervenciones conductuales [158]. Además, el reciente auge de la investigación de efectividad comparativa, especialmente en los Estados Unidos, ha llevado a usar el término “efectividad” más allá de los límites de su significado científico [158]. En este contexto, los mismos autores postulan que corresponde también a los editores de las revistas científicas asegurar el uso correcto y riguroso de la terminología [158].

5.4. Fortalezas, limitaciones e implicaciones

5.4.1 Fortalezas

La principal fortaleza de este trabajo se basa en que aborda la brecha existente entre la investigación y la práctica clínica, un tema de gran interés en la investigación en general, pero particularmente en el campo de las terapias físicas (donde se incluye la terapia manual) tanto por parte de investigadores como de clínicos.

Si bien existe algún trabajo similar en el campo [48,161], este trabajo constituye no sólo un análisis exhaustivo del grado de pragmatismo de los estudios auto-denominados como pragmáticos sino también una evaluación de su reporte y su riesgo de sesgo. Además, se incluyó toda la literatura existente publicada con la etiqueta “pragmático” en este campo.

Otra fortaleza de este trabajo ha sido contar con un amplio equipo multidisciplinar formado por expertos en terapia manual y expertos en metodología que ha permitido abarcar de forma multidisciplinar el análisis de los resultados.

Y en último lugar, pero más importante, que este trabajo de tesis completa y complementa el trabajo iniciado por Gerard Álvarez en su tesis doctoral [92].

5.4.2 Limitaciones

Nuestra revisión tiene varias limitaciones potenciales. Aunque el equipo de revisores incluyó expertos en el campo de la TM y metodólogos expertos, lo que proporcionó un enfoque multidisciplinario para la extracción de datos, esto podría introducir cierta heterogeneidad en el proceso de extracción de datos que podría haber afectado a los resultados. Además, se requiere un conocimiento profundo de cada intervención evaluada para valorar adecuadamente el pragmatismo de un estudio determinado. Otros autores han informado de dificultades similares [32,50,98,112].

Para mitigar el impacto de estas limitaciones, realizamos un workshop con la Dra.

122

Loudon, pilotamos el formulario de extracción de datos y discutimos ampliamente las valoraciones de antemano. Además, emparejamos a todos los revisores con la IP para discutir las puntuaciones y minimizar la subjetividad. Aunque no se requirió la intervención de un tercero para resolver las discrepancias, se necesitaron discusiones frecuentes entre los revisores para concluir una puntuación fiable.

Por último, la muestra incluida en esta revisión podría no considerarse relevante y podría no representar todas las modalidades de TM. Sin embargo, incluyó todos los estudios pragmáticos publicados en el campo de la TM desde su inicio hasta enero de 2024 y se emparejaron los grupos a partir de este número de artículos obtenido para equiparar las comparaciones.

5.4.3 Implicaciones

Tal como se expuso en el apartado de justificación de esta tesis doctoral (punto 1.4), la línea de investigación que aquí culmina representa la continuidad y consolidación de un trabajo iniciado previamente con la tesis doctoral de Gerard Álvarez [89]. Ambas tesis han sido concebidas de manera complementaria, con el propósito común de abordar de forma crítica y constructiva los principales retos metodológicos que plantea la investigación en Terapia Manual, ámbito donde tanto la doctoranda (SR) como su director (GA) se desempeñan profesionalmente.

A lo largo de estos dos proyectos doctorales, se han generado datos empíricos y reflexiones teóricas que permiten caracterizar con mayor precisión el estado actual de la evidencia científica en TM. En concreto, se ha profundizado en el análisis de la calidad metodológica y el reporte de los estudios existentes, se han identificado las principales limitaciones que comprometen su validez y aplicabilidad clínica, y se han planteado criterios para su mejora.

Como resultado de este proceso de investigación acumulativa y reflexión crítica, se

presenta a continuación una propuesta de aportación original a la disciplina: un protocolo para un estudio DELPHI sobre una guía metodológica estructurada, acompañada de un algoritmo de ayuda al diseño de ECAs en TM. Este material pretende sintetizar las recomendaciones más relevantes derivadas tanto de la literatura científica como de la experiencia adquirida durante el desarrollo de ambas tesis, y pretende ofrecer un marco práctico y fundamentado para avanzar en la calidad de la investigación en TM.

5.4.3.1 Propuesta de recomendaciones para DISEÑAR, REPORTAR I MEJORAR LA CALIDAD de los estudios de efectividad en TM: OPTIMaT (Optimizing Practices in Trial design and reporting for Manual Therapy)

Con los datos obtenidos en los estudios incluidos en ambas tesis doctorales [1,80,144,145,162,163] más recomendaciones metodológicas para el diseño de IC [1,6,28,29,32,48,49,54,65,73,77,80,91,131,136,142,144,145,161,162,164,165] adaptadas al campo de la TM se ha podido recoger una serie de recomendaciones para diseñar, reportar y mejorar la calidad de los estudios de efectividad en terapia manual (tabla 6).

Con esta información la doctoranda SR y su director GA han confeccionado un protocolo para un estudio DELPHI donde a partir de una búsqueda sistemática de bibliografía se pueda diseñar una guía centrada en el diseño de ECAs en TM (**Anexo 4**). Esta guía tendría como objetivo guiar a los investigadores en el campo de la TM a diseñar adecuadamente los ensayos clínicos manteniendo una coherencia entre la pregunta de investigación, los objetivos planteados, la metodología y la terminología. Dicha propuesta, constaría de recomendaciones extraídas de la literatura científica agrupadas en 10 fases clave del diseño, implementación y reporte de un estudio experimental. Esta guía de recomendaciones abordaría tanto la fase de diseño como la fase de reporte puesto que la literatura científica en el campo de las terapias físicas revela una clara correlación entre

la calidad de un estudio y la calidad de su reporte [144]. Mejoras en estos dos elementos pretenderán ayudar a incrementar tanto la validez interna como la validez externa de los estudios en el campo de las terapias manuales y aumentar la calidad metodológica en este campo de investigación.

Tabla 6. Recomendaciones para DISEÑAR y REPORTAR los estudios experimentales en Terapia Manual

RECOMENDACIONES	ESTUDIOS DE EFICACIA	ESTUDIOS DE EFECTIVIDAD		
		ECA HÍBRIDO DE EFECTIVIDAD	ECA DE EFECTIVIDAD COMPARATIVA	ECA PRAGMÁTICO
FASE 1. ADECUACIÓN DE LA PREGUNTA DE INVESTIGACIÓN, INTENCIÓN DEL ESTUDIO Y TERMINOLOGÍA				
PREGUNTA DE INVESTIGACIÓN	Investigar el efecto de una intervención en condiciones ideales	Investigar el efecto de una intervención en condiciones parecidas a la práctica real		

		Estudiar efectos fisiológicos en condiciones reales Identificar factores contextuales	Comparar con otra intervención activa	Implementar una intervención
JUSTIFICACIÓN DEL ESTUDIO	La pregunta de investigación debe realizarse de acuerdo con la literatura existente.			
	El estado de la literatura sobre la pregunta de investigación debe justificarse en la sección de introducción.			
	Establecer la justificación del estudio que se pretende realizar a partir de estudios basados en mecanismos y/o de plausibilidad en el caso de querer	Establecer la justificación del estudio que se pretende realizar a partir de estudios previos de eficacia, estudios con control placebo o revisiones sistemáticas.		

	realizar un estudio de eficacia	
INTENCIÓN DEL ESTUDIO	La sección de la introducción del estudio debe contener una descripción exhaustiva de lo que el autor pretende abordar con el estudio para poder interpretar de forma realista los resultados	
	Los ECA de eficacia deben realizarse como primer paso para establecer cómo podría funcionar una intervención en condiciones controladas y detectar efectos específicos y adversos.	Los ECA de efectividad deben realizarse cuando existan suficientes estudios mecanicistas o de eficacia con control placebo sobre una intervención. Deben realizarse en condiciones clínicas reales.

			Cuando el estudio trate de tratamientos de bajo riesgo ampliamente utilizados en entornos clínicos, puede considerarse la posibilidad de comparar su efectividad con otra intervención de uso común, aunque no existan estudios de eficacia controlados con placebo
USO DE TERMINOLOGÍA COHERENTE	Usar el término «eficacia» para los diseños explicativos.	Use el término «efectividad» o «pragmático» para los diseños de efectividad comparativa.	
FASE 2. ELECCIÓN DEL TIPO DE ECA			
	Diseños paralelos		
		Ensayos cruzados, Steped-wedge, Ensayos de preferencia	
			Ensayos por clústeres
			N=1
FASE 3. ELEGIBILIDAD DE LOS PARTICIPANTES, RECLUTAMIENTO Y PROCESO DE ALEATORIZACIÓN			
INCLUSIÓN PARTICIPANTES	Es aconsejable reducir la heterogeneidad que no sea necesaria para responder a la pregunta de investigación.		

	Incluir a aquellas personas que se espera que respondan mejor a la intervención.	Población amplia que varíe en edad, gravedad, comorbilidad y exposición a otras intervenciones.		
	Excluir las comorbilidades y personas de riesgo.	Los criterios de exclusión deben basarse en aquellas contraindicaciones propias de las técnicas de TM, como tumores, uso prolongado de corticosteroides, inestabilidad articular, infecciones (aunque dependerá de cada técnica en particular).		
	Se pueden incluir sujetos sanos	No incluir sujetos sanos		
RECLUTAMIENTO	Se puede usar cualquier método de reclutamiento	usar los propios métodos que se usarían para reclutar pacientes en la práctica clínica habitual.		
ALEATORIZACIÓN	Aleatorización simple	Considerar métodos alternativos a la aleatorización simple: Estudios de preferencia, aleatorización por grupos (cluster randomisation), diseños stepped-wedge, enrichment methods		

	Ocultar la asignación de la secuencia aleatoria tanto en estudios de eficacia como de efectividad.			
FASE 4. ENTORNO Y ORGANIZACIÓN PARA LLEVAR A CABO LA INTERVENCIÓN				
ENTORNO	El estudio se puede realizar en cualquier tipo de entorno, ya sea un entorno clínico, académico y/o centro especializado	Deben llevarse a cabo en entornos y organizaciones que proporcionen la práctica clínica habitual.		
	Estudios unicéntricos	Estudios unicéntricos	Estudios unicéntricos	Preferiblemente estudios multicéntricos
		Cuando el estudio se realice en un solo centro, es muy importante dar detalles sobre el entorno, el sistema sanitario, los procesos reguladores de la intervención y los tipos de pacientes que se incluyen		
		Si los estudios de efectividad se realizan en un centro especializado, indicarse explícitamente en la sección de introducción que los resultados obtenidos se pretenden que sean generalizables para este tipo de entorno y pacientes.		

		Una vez publicados suficientes estudios de efectividad unicéntricos sobre una misma pregunta de investigación, se debería considerar plantear un estudio multicéntrico siempre y cuando sea económicamente viable.
ORGANIZACIÓN INTERVENCIÓN	La experiencia de los terapeutas en la realización de una intervención manual influye significativamente en los resultados. Por lo tanto, aunque esto puede comprometer la generabilidad de los resultados del ensayo, debe tenerse en cuenta tanto en los estudios de eficacia como en los de efectividad. Debe justificarse en la sección de métodos o en la sección de limitaciones	
	Incluir terapeutas con diferentes años de experiencia y formación para ayudar a generalizar los resultados del ensayo. En este caso y si la muestra lo permite, se debería incluir un análisis por subgrupos de terapeutas en el análisis estadístico de los resultados	
		Se debe evitar añadir recursos adicionales en beneficio del ensayo para garantizar la generabilidad de los resultados a diferentes entornos
FASE 5. INTERVENCIONES EXPERIMENTALES Y CONTROL		
INTERVENCIÓN EXPERIMENTAL	Se puede utilizar una única técnica o una intervención protocolizada en	Evitar técnicas protocolizadas o únicas

	estudios de eficacia o mecanicistas que investiguen los efectos neurofisiológicos de la TM			
	Utilizar una intervención no protocolizada y con un enfoque centrado en el paciente en estudios de eficacia con control placebo/sham o estudios de efectividad			
		Las co-intervenciones son habituales en los pacientes que consultan por terapia manual. Deberían permitirse en los estudios de efectividad que pretendan asemejarse a la práctica clínica		
INTERVENCIÓN CONTROL	En intervenciones control con placebo, se debe elegir un placebo fiable			

	Las intervenciones control deben reproducir tantos componentes del tratamiento investigado como sea posible, aparte de los componentes cuyo efecto se pretenda estudiar en el estudio		
	Los estudios de eficacia deben compararse con un placebo [135] y a poder ser con un placebo y un grupo no-tratamiento	Antes de plantear un estudio de efectividad, es imprescindible averiguar si se han realizado estudios de eficacia comparados con placebo	
		Es preferible, cuando sea posible, un diseño de 3 brazos con un grupo de control con otra intervención y otro grupo control con placebo.	
FASE 6. SEGUIMIENTO Y ADHERENCIA A LA INTERVENCIÓN			
SEGUIMIENTO	Sólo deben incluirse sesiones únicas en los	Debería incluirse un seguimiento superior a 3 meses cuando la eficacia de la intervención ya haya sido estudiada en estudios previos de eficacia y efectividad a corto plazo	

	estudios mecanicist as de eficacia	
		Se debe adaptar el seguimiento a la historia natural de la condición patológica estudiada y reportarlo adecuadamente
ADHERENCIA	El seguimient o de la adherencia de los participant es es crucial en un estudio de eficacia	El seguimiento de la adherencia de los participantes en los estudios de efectividad debe ceñirse a las mismas estrategias usadas en la práctica habitual, por ejemplo, confirmar visitas antes de la cita del paciente que es una medida habitual en la práctica rea
FASE 7. VARIABLES DE RESULTADO Y ENMASCARAMIENTO		
VARIABLES DE RESULTADO	Utilizar variables objetivas y/o fisiológicas	Utilizar variables objetivas y/o PROMS
		Utilizar PROMS
		Las PROMs deben evaluarse al realizar estudios de efectividad. Pero también se pueden utilizar en estudios de eficacia juntamente con variables objetivas
		Es aconsejable simplificar la elección de las variables de resultado en los estudios de efectividad, como

		utilizar medidas con pocas escalas (validadas) en lugar de cuestionarios de discapacidad con múltiples preguntas		
		Una vez que un estudio ha probado la eficacia de una intervención sobre un resultado fisiológico, debe evaluarse la relación entre el resultado fisiológico y las PROMs		
		La iniciativa Core Outcome Measures in Effectiveness Trials (http://www.comet-initiative.org) representa una herramienta útil para los investigadores de TM		
ENMASCARAMIENTO	Cegar siempre el evaluador externo tanto en estudios de eficacia como de efectividad			
		Evaluar siempre el éxito del enmascaramiento al final del estudio en cualquier diseño, y también cuando se utilizan PROMs (para garantizar el enmascaramiento del evaluador externo)		
	No utilizar el término doble-cego en estudios donde no se ha cegado el terapeuta, por lo que si se enmascara el asesor externo o estadista, se debería utilizar el término “ciego dual”			

		Utilizar métodos de enmascaramiento cuando los datos de eficacia sean inconsistentes y cuando sea compatible con los objetivos del estudio	
			Es posible utilizar ensayos de efectividad comparativa sin enmascaramiento cuando las intervenciones están bien establecidas en la práctica clínica
		Implementar alternativas al enmascaramiento del participante, como el enmascaramiento de la hipótesis y los diseños de Zelen	
FASE 8. ANÁLISIS ESTADÍSTICO Y RELEVANCIA CLÍNICA			
ANÁLISIS ESTADÍSTICO	Realizar análisis por intención de tratar (ITT) tanto en los estudios de eficacia y efectividad		
	Incluir análisis intra-grupos y entre-grupos		
		Si el estudio va a incluir una cohorte muy heterogénea, el estudio debe diseñarse con un tamaño de muestra suficientemente grande para garantizar un análisis robusto de subgrupos	
RELEVANCIA CLÍNICA	Se debe realizar un cálculo a priori del tamaño de la muestra y un análisis de la potencia		
	Debe tenerse en cuenta las Diferencias mínimas clínicamente importantes (MICD) en lugar del valor estadístico «p»		
	Se aconseja el uso de intervalos de confianza como estimación de precisión para realizar una mejor interpretación clínica de la literatura y establecer la magnitud real del efecto observado		

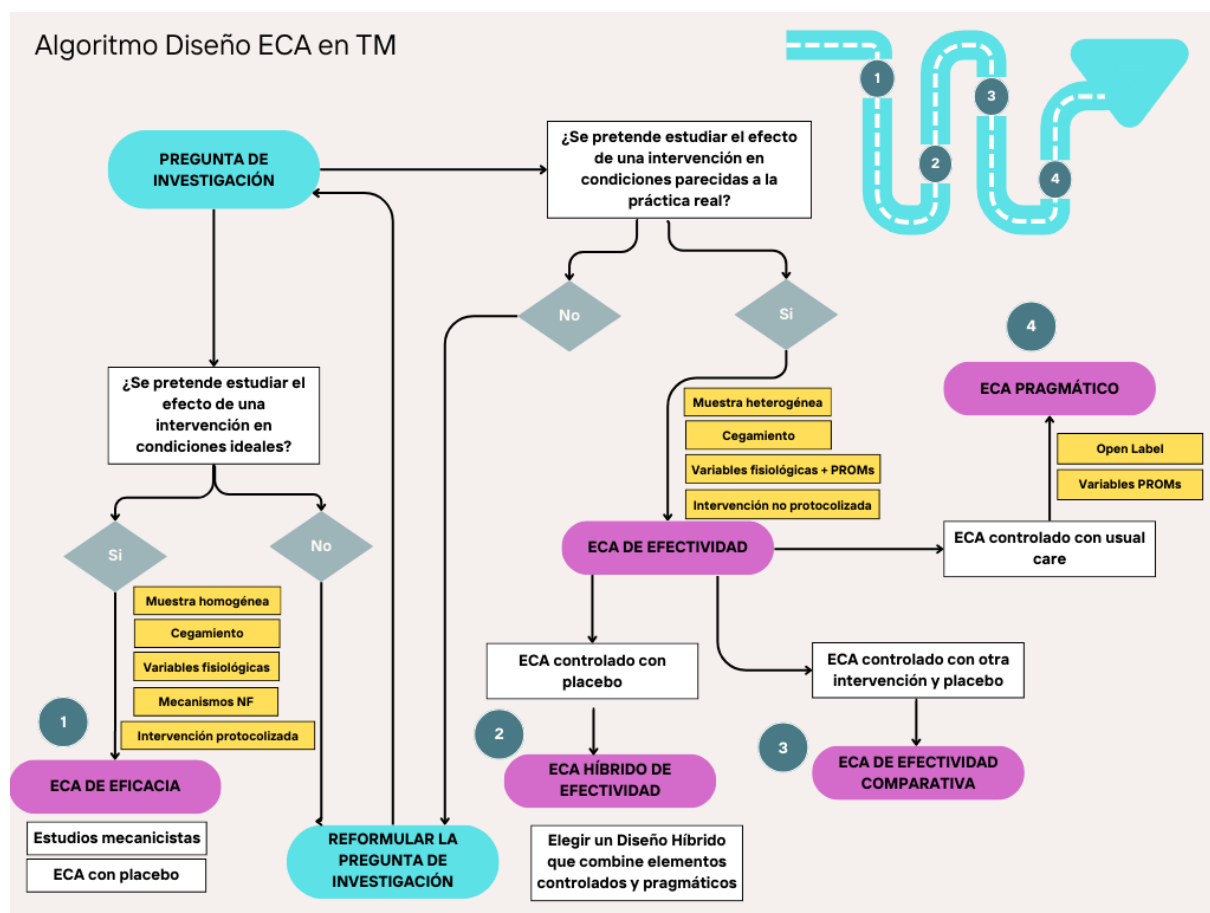
		Debe incluirse la relevancia clínica de los resultados
ANÁLISIS DE SUBGRUPOS		Incluir la experiencia del terapeuta en el análisis de subgrupos
		Incluir el análisis de subgrupos en los estudios de efectividad con poblaciones heterogéneas
		Valorar e incluir las expectativas de los participantes en el análisis de subgrupos
		Analizar, como mínimo, los factores de confusión clínicos más relevantes (por ejemplo, comorbilidades y tratamientos concomitantes)
FASE 9. REGISTRO DEL PROTOCOLO		
Registrar prospectivamente el protocolo del estudio en todos los tipos de ECAs		
FASE 10. REPORTE DEL ESTUDIO		
	Utilizar la extensión CONSORT para el reporte de intervenciones no farmacológicas en ensayos de eficacia.	Utilizar la extensión CONSORT para ensayos pragmáticos e intervenciones no farmacológicas en estudios de efectividad.
		Utilizar la extensión CONSORT para el reporte de PROMs si se utilizan en el estudio.

	Utilice la lista de comprobación TIDIER para reportar adecuadamente la intervención.	
	Utilizar la lista de comprobación TIDIER para reportar la intervención con placebo.	
		Añadir como material suplementario la intervención adaptada realizada a cada paciente y su justificación en los estudios de efectividad para permitir la replicación
	Reportar los métodos utilizados para limitar el sesgo cuando el cegamiento es imposible.	
	Reportar los métodos utilizados para verificar el cegamiento de los participantes.	
	Reportar el uso de las Guías de diseño y reporte utilizadas.	
	Reportar el registro del protocolo del estudio	

5.4.3.2 Propuesta de algoritmo para diseñar y evaluar las Intervenciones Complejas en Terapia Manual a través de un ECA:

La segunda parte del estudio DELPHI pretendería proponer un algoritmo (Figura 9) a partir de las recomendaciones anteriores y la reflexión sobre ellas para desarrollar y evaluar las intervenciones de TM a través de un ECA. Esta propuesta se basaría en 4 fases, por las que, de forma consecutiva, se puede desarrollar una intervención de TM a través de un ECA. En cada una de las fases, se incluye qué características metodológicas deberían tener los estudios en cada fase. Si bien la propuesta de algoritmo se basa en el diseño y la evaluación de la TM a través de un ECA, se propondrían diversos diseños alternativos que podrían complementar cada fase.

Figura 9. Propuesta de algoritmo para desarrollar y evaluar las intervenciones de Terapia Manual



1: Fase 1, 2: Fase 2, 3: Fase 3, 4: Fase 5, ECA: ensayo controlado aleatorio, NF: neurofisiológicos, PROMs (patient reported outcomes)

5.4.3.3 Futuras investigaciones e implicación en entornos académicos

Si bien estas propuestas están basadas en literatura publicada recientemente, son fruto de la reflexión personal de los resultados obtenidos en nuestra tesis doctoral. Es por ello que proponemos varios proyectos en la misma línea de investigación que puedan reforzar y aportar metodología rigurosa a las propuestas realizadas en nuestra tesis. Por ejemplo:

- a. Realizar una scoping review o revisión narrativa sobre las recomendaciones para el diseño y reporte de las intervenciones en TM.
- b. Realizar un Delphi con panel de expertos para establecer la guía de recomendaciones para diseñar y reportar los estudios en TM.
- c. Explorar el conocimiento que tiene los investigadores en TM sobre los estudios pragmáticos para intervenciones no farmacológicas para entender esta falta de aumento a lo largo del tiempo.
- d. Evaluar herramientas de evaluación del riesgo de sesgo en función de la pregunta de investigación y el diseño del estudio.
- e. Incluir, a nivel académico, formación específica sobre los estudios pragmáticos.
- f. Crear y usar bases de datos (medical records and electronic database) para facilitar la inclusión de participantes en estudios a larga escala

6. CONCLUSIONES

- Si bien los ECAs de efectividad en TM han aumentado en los últimos años, no lo han hecho aquellos ECAs autodenominados pragmáticos.
- Los ECAs de efectividad en TM tienen una actitud moderadamente pragmática, independientemente de si se autodenominan pragmáticos o no y de si utilizan un grupo de control con placebo o no. Estos estudios son pragmáticos en cuanto a la elegibilidad de los participantes, el reclutamiento y las medidas de resultado, pero explicativos en cuanto a la intervención y el contexto. La mayoría de ECAs de efectividad pretenden asemejarse a la práctica clínica, pero las características pragmáticas no se muestran dentro del ensayo.
- El reporte de los ECAs de efectividad en TM suele ser inadecuado, siendo mejor en los estudios autodenominados pragmáticos.
- Mientras que el sesgo de selección es menor cuando los estudios se autodenominan pragmáticos, el sesgo de detección es menor cuando se utiliza un control placebo que otra intervención o la «atención habitual».
- En general, ni la calidad del reporte ni el riesgo de sesgo están relacionados con una actitud pragmática mayor o menor. Los ensayos con un RoB bajo han mostrado un mayor cumplimiento de los ítems de la guía de reporte CONSORT.
- Cuando usar los términos eficacia y/o efectividad parece estar poco claro entre la comunidad de investigadores de TM, utilizando el término efectividad en diseños con características explicativas.

7. ABREVIATURAS

TM: TERAPIA MANUAL

IC: INTERVENCIÓN COMPLEJA

ICs: INTERVENCIONES COMPLEJAS

PROMs: PATIENT-REPORTED OUTCOMES MEASURES

ECA: ENSAYO CLÍNICO ALEATORIZADO

ECAs: ENSAYOS CLÍNICOS ALEATORIZADOS

ITT: INTENCIÓN A TRATAR (INTENTION-TO-TREAT)

EP: ENSAYO CLÍNICO ALEATORIZADO DE EFECTIVIDAD AUTODENOMINADO PRAGMÁTICO

EPs: ENSAYOS CLÍNICOS ALEATORIZADOS DE EFECTIVIDAD AUTODENOMINADOS PRAGMÁTICOS

PRECIS: PRAGMATIC-EXPLANATORY CONTINUUM INDICATOR SUMMARY

EE: ENSAYO CLÍNICO ALEATORIZADO DE EFECTIVIDAD NO AUTODENOMINADO PRAGMÁTICO

EEplacebo: ENSAYO CLÍNICO ALEATORIZADO DE EFECTIVIDAD NO AUTODENOMINADO PRAGMÁTICO CON GRUPO CONTROL PLACEBO

EEotrainervención: ENSAYO CLÍNICO ALEATORIZADO DE EFECTIVIDAD NO AUTODENOMINADO PRAGMÁTICO CON OTRA INTERVENCIÓN ACTIVA COMO GRUPO CONTROL

CONSORT: CONSOLIDATED STANDARDS FOR REPORTING TRIALS

ROB: RISK OF BIAS

TIDieR: TEMPLATE FOR INTERVENTION DESCRIPTION AND REPLICATION

8. ANEXOS

8.1 Anexo 1

Estrategia de búsqueda

MEDLINE

- #1 "Musculoskeletal Manipulations"[Mesh]
 - #2 "Osteopathic Physicians"[Mesh]
 - #3 "Osteopathic Medicine"[Mesh]
 - #4 manual therap*[tiab]
 - #5 manipulative therap*[tiab]
 - #6 spinal manipulation*[tiab]
 - #7 manipulation[ti]
 - #8 osteopath*[tiab]
 - #9 chiroprac*[tiab]
 - #10 musculoskeletal manipulation*[tiab]
 - #11 mobilization[ti]
 - #12 mobilisation[ti]
 - #13 motion[tiab]
 - #14 myofascial[ti]
 - #15 massage[ti]
 - #16 neurodynamic technique*[tiab]
 - #17 passive movement*[tiab]
 - #18 passive mobilizat*[tiab]
 - #19 passive mobilisat*[tiab]
 - #20 physical therap*[tiab]
 - #21 thrust technique*[tiab]
- 145

- #22 cranial technique*[tiab]
- #23 soft tissue technique*[tiab]
- #24 connective tissue technique*[tiab]
- #25 strain counterstrain[tiab]
- #26 gliding*[ti]
- #27 passive stretching*[tiab]
- #28 trigger point*[ti]
- #29 biodynamic*[tiab]
- #30 movement dynamic*[tiab]
- #31 #1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR
#12 OR #13 OR #14 OR #15 OR #16 OR #17 OR #18 OR #19 OR #20 OR #21 OR #22
OR #23 OR #24 OR #25 OR #26 OR #27 OR #28 OR #29 OR #30
- #32 "Pragmatic Clinical Trial"[pt]
- #33 "Pragmatic Clinical Trials as Topic"[Mesh]
- #34 #32 OR #33
- #35 #31 AND #34
- #36 #31 NOT #35
- #37 (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])
- #38 #36 AND #37
- #39 pragmatic[tiab]
- #40 naturalistic[tiab]
- #41 comparative effectiveness[tiab]
- #42 effectiveness[ti]
- #43 #39 OR #40 OR #41 OR #42
- #44 #38 AND #43

Cochrane Central Register of Controlled Trials (CENTRAL)

- #1 MeSH descriptor: [Musculoskeletal Manipulations] explode all trees
- #2 MeSH descriptor: [Osteopathic Physicians] explode all trees
- #3 MeSH descriptor: [Osteopathic Medicine] explode all trees
- #4 (manual NEXT therap*):ti,ab
- #5 (manipulative NEXT therap*):ti,ab
- #6 (spinal NEXT manipulation*):ti,ab
- #7 manipulation:ti
- #8 osteopath*:ti,ab
- #9 chiroprac*:ti,ab
- #10 (musculoskeletal NEXT manipulation*):ti,ab
- #11 mobilization:ti
- #12 mobilisation:ti
- #13 motion:ti,ab
- #14 myofascial:ti,ab
- #15 massage:ti
- #16 (neurodynamic NEXT technique*):ti,ab
- #17 (passive NEXT movement*):ti,ab
- #18 (passive NEXT mobilizat*):ti,ab
- #19 (passive NEXT mobilisat*):ti,ab
- #20 (physical NEXT therap*):ti,ab
- #21 (thrust NEXT technique*):ti,ab
- #22 (cranial NEXT technique*):ti,ab
- #23 (soft NEXT tissue NEXT technique*):ti,ab
- #24 (connective NEXT tissue NEXT technique*):ti,ab
- #25 (strain NEXT counterstrain):ti,ab
- #26 gliding*:ti
- #27 (passive NEXT stretching*):ti,ab
- #28 (trigger NEXT point*):ti,ab
- #29 biodynamic*:ti,ab

#30 (movement NEXT dynamic*):ti,ab
#31 #1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR
#12 OR #13 OR #14 OR #15 OR #16 OR #17 OR #18 OR #19 OR #20 OR #21 OR #22
OR #23 OR #24 OR #25 OR #26 OR #27 OR #28 OR #29 OR #30
#32 pragmatic:ti,ab
#33 naturalistic:ti,ab
#34 (comparative NEXT effectiveness):ti,ab
#35 effectiveness:ti
#36 #32 OR #33 OR #34 OR #35
#37 #31 AND #36
#45 #35 OR #44

8.2 Anexo 2

Formulario de extracción de datos

1. BIBLIOMETRIC INFORMATION REVIEWER 1		
Code		
Author		
Title		
Year publication		
Journal		
journal category	TM/no specialist	
Article reference		
b) Study characteristics		
Country of the setting of the study		
Setting	uni/ multicentric/ unclear	

N of participants		
Recruitment period	copy/paste	período que duró el reclutamiento,

Patient's origin	hospitalised/ outpatient/ academic enviroment/ unclear	Outpatient se considerará siempre que la intervención no sea con pacientes hospitalizados (incluye: clínicas, consultas,) Academic enviroment: universidades, centros académicos, o de investigación donde no se realizaría normalmente la intervención
Co-interventions	YES/NO/ UNCLEAR	SI: se permite que los participantes puedan realizar otras terapias/mediación mientras dura el estudio NO: se especifica no realizar ninguna terapia/intervención mientras dura el estudio
Number of sesiones	listado hasta 40 opción:individ ual	Introducir el número de sesiones que se realizan, si es un ytratamiento individualizado y se adapta el número de sesiones a cada paciente escoger opción "individual"
Lenght of complete treatment		Si se realizan 2 sesiones en 2 meses, en una semana o en 1 día (el follow-up se recoge después). si se puede ponerlo en semanas, si es un día poner un día

Additional training of care-givers	YES/NO/ UNCLEAR	<p>Si: se realizó una formación adicional a los profesionales que realizaban la intervención.</p> <p>NO: si no está esta información no se considera additional training si es para garantizar la mejor realización del estudio</p>
Staff expertise	YES/NO/ UNCLEAR	<p>Si: Se reportan los años de experiencia de los profesionales que realizan la intervención como dato importante. o el hecho de tener una formación especializada acorde con la intervención</p> <p>NO: si se interpreta que no es importante la experiencia del personal o se declara que los profesionales tenían diferentes grados de experiencia</p> <p>UNCLEAR: si no se puede interpretar</p>

Extra staff	YES/NO/ UNCLEAR	SI: para realizar la intervención se requiere incorporar más personal que el que se utilizaría en un usual care NO: si aunque no se especifica se entiende que no se aumenta el personal para realizar el estudio UNCLEAR: si no está la información
Nombre intervención control	copy/paste	
Follow up	copy/paste	
Blinding	YES/NO/ UNCLEAR	
Participants		
Therapists		
External assessors		si son distintos al terapeuta
Statistician		si no está reportado es un no
Primary outcome	PROMS/ physiological usual care/ physiological	PROMS: cuestionarios, escalas, todo lo que sea reportado por el paciente Physiological usual care: BP, HR, RR, nistagmus, Physiological external expertise: blood test, glucose, MRI., specific devices

	external expertise/ unclear/mixed	mixed: PROMS y physiological
Statistical analysis (ITT)	YES/NO/ unclear	YES: si se realizó o no se produjeron pérdidas NO: no se realizó ITT
Is there a description of the pragmatic attitude of the article: the intent of the trial	YES/NO/ UNCLEAR	párrafo donde se describa la intención de diseñar un estudio lo más parecido al usual care. puede estar en el abstract/intro/métodos/discusión. Y puede estar tanto en estudios pragmáticos como de efectividad
(Copy and paste)		copiar el fragmento que justifique vuestra decisión
Does the article use any design or reporting tool?	design/ reporting/ both/no	tiene que estar especificado en el texto, o flowchart

Spirit	YES/NO	
Precis or PreciS-2	YES/NO	
Consort	YES/NO	
Tidier	YES/NO	
Does the study have any funding?	YES/NO/ UNCLEAR	si describe que no tiene, responder no unclear si no se reporta

2. INTENT OF THE TRIAL	ALL REVIEWERS
Assessed by using the following three qualitative categories pre-established by the reviewer's team:	I) aimed to clinical informing for decision-making II) aimed to resemble clinical practice, III) aimed both to resemble clinical practice and to clinical inform for decision-making.

3. RATIONALE OF THE INTERVENTION GIVEN BY THE AUTHORS	ALL REVIEWERS
---	---------------

It refers to the proof-of-concept information existing in the literature that the study's author reports answering the research question in the introduction section. Reviewers pre-established 3 types of rationale:	I) report the existence of mechanistic experiments explaining the mechanisms of action of the intervention
	II) report the existence of sham-controlled studies testing the intervention against a placebo,
	III) report comparative effectiveness data trials about the intervention.

4. EXPERIMENTAL AND CONTROL INTERVENTIONS	REVIEWER 1
EXPERIMENTAL INTERVENTION	
Combination of techniques no standardized	a combination of MT techniques chosen by the practitioner based on patient's needs
Isolated technique no standardized	a single MT technique chosen by the therapist based on patient's needs
Combination of techniques protocol	a combination of MT techniques standardised by the trial protocol
Combination of therapies no standardized	a combination of different modalities of treatment based on patient's needs, pe; MT combined with exercises
Isolated technique protocol	a single MT technique standardized by the trial protocol

Combination of therapies protocol	a combination of therapies standardized to all the patients by the trial protocol
CONTROL INTERVENTION	
other intervention	other active intervention
placebo	an inactive intervention that looks like the drug or treatment being tested
usual care	an intervention that is the care the targeted patient population would be expected to receive as part of the normal practice
no intervention	another group that is not receiving any intervention

5. LIMITATIONS REPORTED BY THE AUTHORS	ALL REVIEWERS
The extraction data was based on the 12 limitations categories proposed by Alvarez et al. (Alvarez et al. 2021)	Randomization and/or concealment
	Blinding
	Lost to follow-up/dropout
	Sample Size
	Inadequate control
	Funding
	Type of setting/care providers
	Inadequate subjects
	Study length and follow up
	Intervention
	Outcome measures
	Compromised generalization of the results

6. PRECIS- 2 TOOL ASSESSMENT: consists of 9 domains, including eligibility, recruitment, setting, organisation, flexibility (delivery), flexibility (adherence), follow-up, primary outcome, and primary analysis. Mean values were calculated, including all reported domains. For this study, reviewers rated the non- reported domains " blank " and excluded them from score. Score of three will be used when a domain is as pragmatic as explanatory.	SCORE (1-5)
DOMAIN 1: ELIGIBILITY	<p>los criterios de inclusión deben incluir aquellos participantes relacionados con la pregunta de investigación y con el primary outcome</p> <p>TENER en cuenta también si el sistema sanitario que se plantea se parece a la realidad o si usa muchos tests para seleccionar</p>
DOMAIN 2: RECRUITMENT	<p>penalizar cuando se realicen esfuerzos para reclutar pacientes de forma distinta a la práctica habitual. Altamente pragmático cuando se recluta de más de un sitio</p>

DOMAIN 3: SETTING	<p>aunque haya varios centros si no hay más de un país no poner un 5, a no ser que el objetivo del estudio sea ver los efectos de la intervención específicamente en un país, o la intervención es fácilmente aplicable en otros países (tener en cuenta situación económica y sanitaria). También si se compara con usual care este puede variar entre países.</p>
DOMAIN 4: ORGANIZATION	<p>Si la formación adicional se realiza para garantizar el funcionamiento del estudio o se considera parte de la intervención no se considerará explicativo. Penalizar el caso que la formación adicional sea para mejorar los efectos de la intervención</p>

DOMAIN 5: FLEXIBILITY DELIVERY	si hay un protocolo ver hasta qué punto se parece a la práctica habitual
DOMAIN 6: FLEXIBILITY ADHERENCE	Aunque habitualmente no se raporta, a veces se puede intuir mirando las justificaciones de las pérdidas en el flowchart
DOMAIN 7: FOLLOW UP	penalizan follow-up muy cortos, con más sesiones de lo habitual o con mucha recolección de datos y/o más personal para realizarlo
DOMAIN 8: PRIMARY OUTCOME	aunque no sea PROMS, puede ser relevante para el paciente (no penalizar)
DOMAIN 9: PRIMARY ANALYSIS	Aunque el texto no especifique si hay ITT o no revisar el flow chart o las tablas de resultados para entender como se han tratado las pérdidas
REVIEWER OPINION QUESTIONS: opinión personal del REVIEWER (subjetiva), se ha añadido una pregunta más	
Is there anything that suggests that the trial not resemble usual clinical practice (copy and paste)	YES/NO/ UNCLEAR
are the results applicable to other usual care settings	YES/NO/

	UNCLEAR
If not, how could you improve the similarity of the study with usual clinical practice?	write

7. CONSORT 2010 ASSESSMENT: all items from the extension for pragmatic trials and non-pharmacological interventions were included ([Moher et al. 2010](#); [Zwarenstein et al. 2008](#); [Boutron et al. 2017](#)).

Se ha añadido la opción not applicable en algunas preguntas. los ítems con ** solo son aplicables si el estudio se autodenomina pragmático. Los ítems * son para todos los artículos ya que es la extensión para intervenciones no- farmacológicas

Title and abstract
1a. Identification as a randomized trial in the title
1b. Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)
Introduction
2a. Scientific background and explanation of the rationale **Describe the health or health service problem that the intervention is intended to address and other interventions that may commonly be aimed at this problem
2b. Specific objectives or hypotheses
Methods
Trial design
3a. Description of trial design (such as parallel, factorial) including allocation ratio *When applicable, how care providers were allocated to each trial group
3b. Important changes to methods after trial commencement (such as eligibility criteria), with reasons
Participants

4a. Eligibility criteria for participants

*When applicable, eligibility criteria for centres and for care providers

**Eligibility criteria should be explicitly framed to show the degree to which they include typical participants and/or, where applicable, typical providers (eg, nurses), institutions (eg, hospitals), communities (or localities eg, towns) and settings of care (eg, different healthcare financing systems)

4b.Settings and locations where the data were collected
Interventions
<p>5a.The interventions for each group with sufficient details to allow replication, including how and when they were actually administered</p> <p>*Precise details of both the experimental treatment and comparator</p> <p>**Describe the comparator in similar detail to the intervention</p>
<p>5b.*Description of the different components of the interventions and, when applicable, description of the procedure for tailoring the interventions to individual participants.</p>
5c.*Details of whether and how the interventions were standardized.
<p>5d.**Describe extra resources added to (or resources removed from) usual settings in order to implement the intervention. Indicate if efforts were made to standardise the intervention or if the intervention and its delivery were allowed to vary between participants, practitioners, or study sites</p>
5e.*Details of whether and how adherence of care providers to the protocol was assessed or enhanced

5f.*Details of whether and how adherence of participants to interventions was assessed or enhanced

Outcomes

6a. Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed **Explain why the chosen outcomes and, when relevant, the length of follow-up are considered important to those who will use the results of the trial
6b. Any changes to trial outcomes after the trial commenced, with reasons
Sample Size
7a. How sample size was determined **If calculated using the smallest difference considered important by the target decision-maker audience (the minimally important difference) then report where this difference was obtained
7b. When applicable, explanation of any interim analyses and stopping guidelines
Randomization
Sequence generation
8a. The method used to generate the random allocation sequence
8b. Type of randomisation; details of any restriction (such as blocking and block size)
Allocation concealment
9. The mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned

Implementation
10. Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions
Blinding

<p>11a.If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how</p> <p>*If done, who was blinded after assignment to interventions (e.g., participants, care providers, those administering co-interventions, those assessing outcomes) and how</p> <p>**If blinding was not done, or was not possible, explain why</p>
11b.If relevant, a description of the similarity of interventions
11c.*If blinding was not possible, description of any attempts to limit bias
Statistical methods
<p>12a.Statistical methods used to compare groups for primary and secondary outcomes</p> <p>*When applicable, details of whether and how the clustering by care providers or centres was addressed</p>
12b. Methods for additional analyses, such as subgroup analyses and adjusted analyses
Results
Participants flow (a diagram is strongly recommended)

13a. For each group, the numbers of participants who were randomly assigned, received intended treatment and were analyzed for the primary outcome
*The number of care providers or centres performing the intervention in each group and the number of patients treated by each care provider or in each centre
**The number of participants or units approached to take part in the trial, the number which were eligible, and reasons for non- participation should be reported
13b. For each group, losses and exclusions after randomization, together with reasons
13c.*For each group, the delay between randomization and the initiation of the intervention
13d. Details of the experimental treatment and comparator as they were implemented
Recruitment
14a. Dates defining the periods of recruitment and follow-up
14b. Why the trial ended or was stopped
Baseline data

15.A table showing baseline demographic and clinical characteristics for each group

*When applicable, a description of care providers (case volume, qualification, expertise, etc.) and centres (volume) in each group

Numbers analysed

16. For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups
Outcomes and estimation
17a. For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)
17b. For binary outcomes, the presentation of both absolute and relative effect sizes is recommended
Ancillary analyses
18. Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing prespecified from exploratory
Harms
19. All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)
Discussion
20. Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses *In addition, take into account the choice of the comparator, lack of or partial blinding, and unequal expertise of care providers or centres in each group

21. Generalizability (external validity, applicability) of the trial findings
*Generalizability (external validity) of the trial findings according to the intervention, comparators, patients, and care providers and centres involved in the trial
**Describe key aspects of the setting which determined the trial results. Discuss possible differences in other settings where clinical traditions, health service organisation, staffing, or resources may vary from those of the trial
22. Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence
23.Registration number and name of trial registry
24. Where the full trial protocol can be accessed, if available
25.Sources of funding and other support (such as supply of drugs), role of funders

8. RISK OF BIAS Cochrane tool (Higgins et al. 2011)	HIGH/ LOW/ UNCLEAR
Selection bias	
Random sequence generation	
Allocation concealment	

Performance bias Blinding participants	
Performance bias Blinding personnel	
Detection bias Blinding (outcome assessment)	
Attrition bias Incomplete outcome data	

Reporting bias	Puntuación: LOW: el artículo informa sobre el registro del protocolo y los métodos son consistentes con los resultados HIGH: no está el protocolo y los resultados no son consistentes. o está el protocolo pero los métodos del paper no son consistentes con los resultados. UNCLEAR: los métodos son consistentes pero no está el protocolo
----------------	---

9. GENERALISABILITY AND APPLICABILITY

we used the model proposed by Murad et al. to assess the generalisability and applicability of the results by combining information related to sample size and PRECIS-2 score ([Murad et al. 2018](#)). It was considered generalisability when having a sample > 100 and recruitment scores ≥ 4 in PRECIS-2 Domain 2 (recruitment). The trial was considered to have applicability of the intervention if scored ≥ 4 in PRECIS-2 Domains 1 and 3 (eligibility and setting).

8.3 Anexo 3

Propuesta de guía para diseñar y reportar ECAs en Terapia Manual, Propuesta de algoritmo para diseñar ECAs de Terapia Manual y Propuesta de protocolo DELPHI

PROPUESTA DE GUÍA PARA DISEÑAR Y REPORTAR ECAs en Terapia Manual

FASE 1. ADECUACIÓN DE LA PREGUNTA DE INVESTIGACIÓN, INTENCIÓN DEL ESTUDIO, TERMINOLOGÍA

El objetivo de la FASE 1 es formular la pregunta de investigación con el objetivo de determinar de forma adecuada la intención del estudio y su metodología [45,48,49]. La pregunta de investigación debe formularse tomando como base la literatura existente [142,145] y por consiguiente, el estudio se diseñará en función de la utilidad que se pretenden dar a sus resultados [142].

Para diseñar un ECA y situarlo correctamente en el continuo eficacia-efectividad deben tenerse en cuenta dos aspectos fundamentales.

El primer aspecto es la justificación que da el autor del estudio para seleccionar un tipo de diseño de estudio específico. En este sentido, es necesario establecer la justificación del estudio que se pretende realizar a partir de estudios basados en mecanismos y/o de plausibilidad en el caso de querer realizar un estudio de eficacia; y basados en estudios previos de eficacia, estudios con control placebo/sham o revisiones sistemáticas en el caso de querer realizar un estudio de efectividad [142,145].

El segundo aspecto es como se reporta la intención que el autor pretende tener mediante

el estudio. Una descripción exhaustiva de lo que el autor pretende abordar con el estudio debería incluirse en la sección de introducción del estudio para poder interpretar de forma realista los resultados [145]. Como norma general, los ECAs de eficacia deberían realizarse como primer paso para establecer cómo podría funcionar una intervención en condiciones controladas y detectar efectos específicos y adversos [6,32,54,142]. Por otro lado, los ECAs de efectividad deberían realizarse cuando existan suficientes estudios mecanicistas o de eficacia con control placebo sobre una intervención, y deberían realizarse en condiciones clínicas reales [54,145]. No obstante, cuando el estudio trate de tratamientos de bajo riesgo ampliamente utilizados en entornos clínicos, puede considerarse la posibilidad de comparar su efectividad con otra intervención usada de forma común en la práctica clínica (estudios de efectividad comparada) [142].

La pregunta de investigación y la intención del estudio determinarán la terminología correcta a utilizar y esta deberá ser clara tanto en el título como en el conjunto del manuscrito. Se debe usar el término «eficacia» para los diseños explicativos (probar la eficacia de una intervención en condiciones ideales)[163]. Por el contrario, se debería usar el término «efectividad» para los diseños de efectividad comparativa y pragmáticos (probar la efectividad de una intervención en condiciones similares a la práctica clínica) [163] .

FASE 2. ELECCIÓN DEL TIPO DE ECA

Tradicionalmente, los ensayos clínicos aleatorizados han priorizado el estudio de la eficacia y la efectividad de las intervenciones. Para ello mayormente se utiliza el clásico diseño de brazos paralelos. Sin embargo, como intervención compleja que es, la TM comprende otros factores que impactan en los resultados. Por esta razón, podría ser recomendable elegir un diseño alternativo que capture y de información acerca de estos aspectos. En el diseño de un ECA se pueden incorporar variantes y adaptaciones metodológicas que incorporen la evaluación de otros aspectos más allá de la eficacia y/o efectividad. Por ejemplo, los ensayos de preferencia permiten incorporar las preferencias

de los pacientes de modo que se pueda evaluar el impacto de estas sobre los resultados y también sobre la adherencia o discontinuidad en el estudio. De manera similar, los ensayos de cuña escalonada (stepped-wedge trial) evalúan la intervención en distintos tiempos temporales dentro del mismo ensayo posibilitando la comparación de los efectos (en distintos grupos o centros) en relación al tiempo de exposición a la intervención. En definitiva, mediante la selección del tipo de ECA, los investigadores pueden incorporar la evaluación de elementos relevantes en el campo de la TM para la evaluación de su importancia en el efecto final observado.

También la elección de un tipo alternativo de ECA puede ser adecuado en estudios de muestras pequeñas o con problemas para la aleatorización, para determinar efectos adversos o estudiar casos raros, o para valorar efectos fisiológicos en situaciones con alta variabilidad interpersonal.

Tipos de ECA alternativos

Ensayos de grupos paralelos: en los ensayos de grupos paralelos, los participantes se asignan al azar a uno de dos o más grupos de tratamiento, y los grupos se siguen y comparan a lo largo del tiempo. Este es el tipo más común de ECA.

Ensayos cruzados (crossover): en los ensayos cruzados, cada participante recibe ambos tratamientos en orden aleatorio, con un período de lavado entre tratamientos para eliminar cualquier efecto persistente del primer tratamiento. Este tipo de ensayo es útil cuando la condición en estudio es crónica y estable en el tiempo.

Ensayos factoriales: en los ensayos factoriales, los participantes se asignan al azar a uno de cuatro grupos de tratamiento, cada uno de los cuales recibe una combinación diferente de dos tratamientos o intervenciones. Este tipo de ensayo permite a los investigadores evaluar los efectos de cada tratamiento individualmente, así como la interacción entre los dos tratamientos.

Ensayos aleatorizados por conglomerados (cluster): En los ensayos aleatorizados por conglomerados, los grupos o conglomerados de personas, como hogares o escuelas, se aleatorizan para recibir diferentes tratamientos o intervenciones. Este tipo de ensayo es útil cuando es difícil o poco práctico aleatorizar a los individuos dentro de los grupos.

Ensayo clínico pragmático: Un ensayo clínico pragmático es un estudio diseñado para

evaluar la efectividad de una intervención en condiciones de práctica clínica habitual, con el objetivo de guiar la toma de decisiones clínicas. Se centra en la evaluación de resultados clínicos importantes para los pacientes y puede incluir poblaciones más diversas y menos controladas que los ECAs tradicionales

Ensayos de cuña escalonada (Stepped-wedge): En los ensayos de cuña escalonada, todos los participantes finalmente reciben la intervención, pero el momento en que la reciben es aleatorio. Este tipo de ensayo es útil cuando se espera que la intervención tenga un efecto positivo y no es ético ocultárselo a ningún participante.

Ensayos adaptativos: En los ensayos adaptativos, el régimen de tratamiento se modifica en función de los resultados observados durante el ensayo. Este tipo de ensayo permite a los investigadores adaptar el tratamiento para optimizar el resultado mientras el ensayo aún está en curso.

Ensayos de preferencia: Un ensayo de preferencia es un tipo de ensayo clínico en el que los participantes tienen la oportunidad de elegir qué tratamiento o intervención preferirían recibir.

Ensayos N=1: Es un tipo de ensayo clínico que involucra a un solo participante, en lugar de un grupo de participantes. También se conoce como diseño experimental de un solo sujeto o de un solo caso. Los ensayos N=1 pueden ser particularmente útiles para evaluar la efectividad de las intervenciones para personas con afecciones raras o complejas

FASE 3. ELEGIBILIDAD DE LOS PARTICIPANTES, RECLUTAMIENTO Y PROCESO DE ALEATORIZACIÓN

Los participantes incluidos en un estudio y la forma en que se incorporan a la investigación pueden influir significativamente en los resultados y ayudar a minimizar la posibilidad de sesgos relacionados con la inclusión. Es necesario determinar, de forma explícita, la población a la que se quiere generalizar los resultados. Los estudios de eficacia deben incluir a aquellas personas que se espera que respondan mejor a la intervención, y se pueden excluir sujetos con comorbilidades y personas de riesgo [142]. Por lo contrario, los estudios de efectividad deben contar con una población amplia que

varíe en edad, gravedad, presencia de comorbilidades y exposición a otras intervenciones [48,80,142,161]. Los criterios de exclusión deberían basarse en aquellas contraindicaciones propias de las técnicas de TM, como tumores, uso prolongado de corticosteroides, inestabilidad articular, infecciones (aunque dependerá de cada técnica en particular) [80,161]. No obstante, en cualquiera de los casos, es aconsejable establecer estrategias para reducir la heterogeneidad que no sea necesaria para responder a la pregunta de investigación [142]

La inclusión de participantes sanos, sólo deberían contemplarse en ensayos mecanicistas que pretendan informar sobre un efecto fisiológico [80].

Para los estudios de eficacia, puede emplearse cualquier método para reclutar participantes[142]. En cambio, para los estudios de efectividad se deben usar los propios métodos que se usarían para incluir pacientes en la práctica clínica habitual [142].

La aleatorización es un elemento clave en cualquier ECA. En los estudios de eficacia, la recomendación está en el uso de aleatorización simple [142]. En los estudios de efectividad, se aconseja considerar métodos alternativos a la aleatorización simple, como la aleatorización por grupos (cluster randomisation) [48,142].

La ocultación de la asignación es un elemento clave para garantizar la efectividad de la aleatorización y evitar el sesgo de selección, y debe realizarse tanto en los estudios de eficacia como en los de efectividad [129].

FASE 4. ENTORNO Y PERFIL DE LOS TERAPEUTAS Y NATURALEZA DE LA INTERVENCIÓN

El dónde y el cómo se lleva a cabo la intervención tendrá un impacto en la capacidad del estudio de generalizar sus resultados. Algunas características específicas, como usar muchos recursos extras para facilitar la investigación o delimitar la inclusión de un determinado perfil de terapeutas, se consideran características muy explicativas porque están alejadas de la práctica habitual. Sin embargo, en el ámbito de la MT, existen

matices que deben tenerse en cuenta.

Los estudios de efectividad deben llevarse a cabo en entornos y organizaciones que proporcionen la práctica clínica habitual [142]. Cuando estos estudios se realicen en un solo centro, es muy importante dar detalles sobre el entorno, el sistema sanitario, los procesos reguladores de la intervención y los tipos de pacientes que se incluyen [32,144,145]. Si estos se realizan en un centro especializado, debería indicarse explícitamente, en la sección de la introducción, que los resultados obtenidos se pretenden que sean generalizables para este tipo de entorno y pacientes [32,144,145]. No obstante, una vez publicados suficientes estudios de efectividad unicéntricos sobre una misma pregunta de investigación, se debería considerar plantear un estudio multicéntrico siempre y cuando sea económicamente viable [145,161]. Es importante remarcar que, especialmente en los estudios de efectividad, debe evitarse añadir recursos adicionales en beneficio del ensayo para garantizar la generabilidad de los resultados a diferentes entornos [142].

El perfil de los terapeutas encargados de la realización de una intervención manual influye significativamente en los resultados [77,145]. Por lo tanto, aunque esto impacta en la generabilidad de los resultados del ensayo, debe tenerse en cuenta tanto en los estudios de eficacia como en los de efectividad. Si bien en los estudios de eficacia es importante reportar explícitamente la experiencia de los terapeutas, se pueden incluir solo terapeutas con amplia experiencia para maximizar los resultados de la intervención. En cambio, en los estudios de efectividad se debería intentar incluir terapeutas con diferentes años de experiencia y formación para ayudar a generalizar los resultados del ensayo. En este caso y si la muestra lo permite, se debería incluir un análisis por subgrupos de terapeutas en el análisis estadístico de los resultados [77]. La información sobre el perfil de los terapeutas debe incluirse en la sección de métodos y/o en la sección de limitaciones [77,145]

FASE 5. INTERVENCIONES EXPERIMENTALES Y CONTROL

Las intervenciones experimental y control son aspectos fundamentales en el diseño de cualquier ECA pero especialmente en los ECAs de TM. Las características específicas de estas intervenciones presentan retos significativos para la investigación evidenciando un conflicto entre los requerimientos metodológicos y la necesidad de evaluar las intervenciones tal y como se aplican en la práctica real. Por lo tanto, es de suma importancia abordar estos aspectos con precisión para garantizar un diseño de investigación sólido que además mantenga el sentido clínico. Esta fase vendrá profundamente determinada por la pregunta de investigación y su intención (FASE 1). Si el objetivo es la evaluación de los efectos neurofisiológicos de la TM (estudios mecanicistas), el establecimiento de una intervención protocolizada es razonable y aconsejable [145]. Sin embargo, si el objetivo es evaluar la efectividad de una intervención manual, el uso de protocolos estandarizados aleja el procedimiento de la práctica habitual [80,145]. Esto es así incluso para estudios de eficacia con control placebo/sham [80,142,145,161]

Por otro lado, las co-intervenciones pueden estar limitadas en los estudios de eficacia pero son una característica habitual en los pacientes que consultan por terapia manual. Por este motivo, deberían permitirse (aunque pueden monitorizarse) en los estudios de efectividad que pretendan asemejarse a la práctica clínica [142,145].

La elección de la intervención de control es tan relevante como la experimental puesto que va a determinar el sentido del experimento. Los estudios de eficacia, al pretender evaluar el efecto neto de una intervención, deberían siempre compararse con un placebo/sham [136] y, si se dispone de la muestra suficiente, con un placebo/sham y un grupo no-tratamiento [54,77]. Ello permite valorar tanto el efecto de la intervención experimental vs el placebo/sham, así como el potencial efecto indirecto del placebo/sham. Una estrategia para minimizar los requerimientos muestrales en el diseño de ensayos con 3 grupos es establecer una asignación 2:2:1. Aunque el grupo no-tratamiento acabe siendo menos numeroso, aplicando las correctas técnicas

estadísticas, se puede conseguir valorar el potencial efecto indirecto de la intervención control.

Aunque el uso de placebos/shams en TM es altamente complicado, existen recomendaciones en la literatura para diseñar un placebo/sham fiable [25,54,77,80,91,164]. Entre otros elementos, un placebo/sham en TM debe contemplar e incorporar todos los elementos procedimentales del tratamiento investigado a excepción del componente/s cuyo efecto se pretenda investigar [25,54]. Particularmente, d'Allessandro et al. proponen el concepto del "Touch Equality-Assumption", basado en igualar los factores no específicos relacionados con el tacto entre grupo experimental y control placebo/sham. Dado que el contacto manual genera efectos placebo y desencadena respuestas psiconeuroinmunoendocrinas, este concepto exige que todas las intervenciones (reales o simuladas) mantengan la misma "coreografía" del tacto: mismas áreas corporales, estilo y tipo de tratamiento (tacto ligero o thrust, papel activo/pasivo del paciente). Para desarrollar procedimientos simulados válidos, es crucial identificar y reportar las características específicas de la técnica probada, garantizando así la credibilidad del tratamiento simulado [25].

Como se ha comentado, en el caso de los estudios de efectividad, es aconsejable averiguar previamente si se han realizado estudios de eficacia comparados con placebo/sham [145], puesto que refuerzan la justificación de una valoración más aplicada. Si esta premisa se da, se pueden valorar otro tipo de controles que incluyen tanto intervenciones de comparación (efectividad comparada) como también intervenciones simuladas en el contexto de la práctica habitual y aplicada. De nuevo, si es posible en términos de potencia estadística, un diseño de 3 brazos con un grupo de control con otra intervención y otro grupo control con placebo [54,77,136,142,145,153] sería aconsejable.

FASE 6. SEGUIMIENTO Y ADHERENCIA A LA INTERVENCIÓN

En el ámbito de la terapia manual, la forma de realizar el seguimiento y de controlar la adhesión de los pacientes al tratamiento varía en función de la profesión que utilice la terapia manual y de los métodos habituales de cada profesional.

En los estudios de efectividad, sugerimos un seguimiento superior a 3 meses cuando la eficacia de la intervención ya haya sido estudiada en estudios previos de eficacia y efectividad a corto plazo [80]. Específicamente en el campo de la TM, es necesario asegurar un seguimiento a largo plazo en las entidades crónicas [166]. Se debería también, adaptar el seguimiento a la historia natural de la condición patológica estudiada y reportarlo adecuadamente [32,80]. Es preferible, sólo incluir sesiones únicas en los estudios mecanicistas de eficacia (estudios pre-post) [48].

El seguimiento de la adherencia de los participantes es crucial en un estudio de eficacia [77]. No obstante, en los estudios de efectividad debe ceñirse a las mismas estrategias usadas en la práctica habitual, por ejemplo, confirmar visitas antes de la cita del paciente que es una medida habitual en la práctica real [142].

FASE 7. MEDIDAS DE RESULTADO Y enmascaramiento

De nuevo, la pregunta de investigación, así como la intención del estudio son los elementos determinantes en la elección de las medidas de resultado.

Como norma general, los estudios de eficacia, diseñados para obtener una valoración objetiva y neta del efecto de una intervención, precisan de medidas de resultado objetivas y no sujetas a sesgos interpretativos por parte de los pacientes o evaluadores [46]. Sin embargo, en los estudios de efectividad, el uso de medidas auto-reportadas por los pacientes (PROMs) es altamente aconsejable. De hecho, la utilización de PROMs se recomienda especialmente para la evaluación de IC [167–169] y responde también al objetivo de captar el input del paciente en los estudios de investigación [170,171]. En cualquier caso, hay situaciones donde la combinación de distinto tipo de variables es

posible e incluso deseable. Por ejemplo, el uso de PROMs puede utilizarse en estudios de eficacia juntamente con variables objetivas [54,77,80]. Una vez que un estudio ha probado la eficacia de una intervención sobre un resultado fisiológico, es aconsejable evaluar la potencial relación entre el resultado fisiológico y las PROMs [77,80]. A pesar de estas posibilidades, es aconsejable simplificar la elección de las variables de resultado especialmente en los estudios de efectividad, por lo que se aconseja utilizar escalas simples (validadas) en lugar de cuestionarios de discapacidad con múltiples preguntas [142].

Otro elemento importante a considerar en relación a la elección de medidas de resultado en los ECAs es su potencial utilización en la investigación secundaria. En este sentido, la utilización de paquetes de medidas validadas para una condición clínica concreta, contribuyen a una mejor homogeneidad de los resultados de la investigación primaria para su uso en productos de síntesis. A este respecto, la iniciativa Core Outcome Measures in Effectiveness Trials (<http://www.comet-initiative.org>) representa una herramienta útil para los investigadores de TM [80].

El enmascaramiento constituye un reto metodológico en el contexto de los estudios de terapia manual con un alto impacto en su validez interna. No obstante, pueden tenerse en cuenta una serie de recomendaciones para mejorar este aspecto del diseño

Se aconseja utilizar métodos de enmascaramiento cuando los datos de eficacia sean inconsistentes y cuando sea compatible con los objetivos del estudio [142]. Como primera recomendación, se aconseja asegurar el enmascaramiento del evaluador externo (tanto en estudios de eficacia como de efectividad) siempre y cuando sea posible [54,77]. A este respecto, no se recomienda utilizar el término doble-ciego en estudios donde no ha sido posible cegar al terapeuta, por lo que, si se enmascara el asesor externo o estadista, se debería utilizar el término “ciego dual” [78]. Sin embargo, incluso el enmascaramiento del evaluador externo puede no ser factible en el caso de que las medidas de resultado sean PROMS (muy comunes en TM y especialmente

recomendados en los ensayos pragmáticos). En este caso, los evaluadores son los propios pacientes a los que no se ha podido cegar, invalidando de facto el enmascaramiento de los evaluadores. De hecho, la literatura informa de que, sin un enmascaramiento adecuado, los estudios con PROMs tienden a mostrar tamaños de efecto mayores [54,77,136,142,145,153]. Por ello, tanto en este caso, como en cualquier circunstancia donde se perciba un posible sesgo de detección (fallo de la estrategia de enmascaramiento) es aconsejable tanto una evaluación post-hoc del éxito del cegamiento como testar la estrategia de enmascaramiento en la fase piloto del estudio. [77,80,144].

Por otro lado, es posible utilizar ensayos de efectividad comparativa sin enmascaramiento cuando las intervenciones están bien establecidas en la práctica clínica [142]. O bien, Implementar alternativas al enmascaramiento del participante, como el enmascaramiento de la hipótesis y los diseños de Zelen [80,142].

FASE 8. ANÁLISIS ESTADÍSTICO Y RELEVANCIA CLÍNICA

Un análisis estadístico adecuado es esencial para considerar que los resultados de un estudio tienen validez interna. Para ello es necesario que la pregunta de investigación, el diseño de la investigación y la base de datos que recoge las variables que se van a analizar tengan coherencia entre ellos.

Se debe realizar análisis por intención de tratar (ITT) tanto en los estudios de eficacia y efectividad [32,49,142]. Además, Incluir análisis intra-grupos y entre grupos [142]. Si el estudio va a incluir una cohorte muy heterogénea, el estudio debe diseñarse con un tamaño de muestra suficientemente grande para garantizar un análisis robusto de subgrupos [142].

Se debe realizar siempre un cálculo a priori del tamaño de la muestra y un análisis de la potencia [80]. Se aconseja el uso de intervalos de confianza como estimación de

precisión para realizar una mejor interpretación clínica de la literatura y establecer la magnitud real del efecto observado. En los estudios de terapia manual deben tenerse especialmente las diferencias mínimas clínicamente importantes (MICD) en lugar del valor estadístico «p» [32] y reportar la relevancia clínica de los resultados en los estudios de efectividad.

En estudios de efectividad con poblaciones heterogéneas se debe incluir un análisis de subgrupos [142]. Además, es aconsejable en el análisis de subgrupos, valorar e incluir las expectativas de los participantes [142] y la experiencia del terapeuta [77,145]. y también, analizar, como mínimo, los factores de confusión clínicos más relevantes (por ejemplo, comorbilidades y tratamientos concomitantes) [142].

FASE 9. REGISTRO DEL PROTOCOLO

El registro prospectivo del protocolo es un indicador de transparencia y buenas prácticas y necesario para garantizar el mantenimiento de la calidad del estudio [144,165].

FASE 10. RECOMENDACIONES EN EL REPORTE

La literatura revela una clara correlación entre la calidad de un estudio y la calidad de su reporte [144]. Es por ello que recomienda imperativamente el uso de las diferentes guías existentes para el reporte de los ECAs en terapia manual.

Se debe utilizar la extensión CONSORT para intervenciones no farmacológicas en ensayos de eficacia [127,128,144] y la extensión CONSORT para ensayos pragmáticos e intervenciones no farmacológicas en estudios de efectividad [97,144]. Para el reporte de intervenciones control con placebo utilizar la extensión CONSORT para intervenciones control placebo y la extensión CONSORT específica para el reporte de PROMs si se utilizan en el estudio [93]. Utilizar la lista de comprobación TIDIER para reportar adecuadamente la intervención [131,141,144] así como TIDIER para intervenciones

placebo, en el caso [172].

Es importante añadir como material suplementario la intervención adaptada realizada a cada paciente y su justificación en los estudios de efectividad para permitir la replicación [144].

Si bien diversos ítems están incluidos en las guías mencionadas anteriormente, es especialmente importante en los estudios de terapia manual reportar los métodos utilizados para limitar el sesgo cuando el enmascaramiento es imposible, reportar los métodos utilizados para verificar el enmascaramiento de los participantes, reportar en el estudio el uso de la Guías de diseño y reporte utilizadas y reportar el registro del protocolo del estudio [144]. Particularmente estos ítems se han tenido poco en cuenta en los ECAs publicados hasta la fecha en el campo de la terapia manual, influenciando la calidad de ellos.

PROPUESTA DE ALGORITMO PARA DISEÑAR Y EVALUAR LAS IC EN TM A TRAVÉS DE UN ECA

A partir de las recomendaciones anteriores y la reflexión sobre ellas, se propone un algoritmo para desarrollar y evaluar las intervenciones de TM a través de un ECA. Esta propuesta se basa en 4 fases, por las que, de forma consecutiva, se puede desarrollar una intervención de TM a través de un ECA donde en cada una de ellas, se incluye qué características metodológicas deberían tener los estudios en cada fase. Si bien la propuesta de algoritmo se basa en el diseño y la evaluación de la TM a través de un ECA, se proponen diversos diseños alternativos que pueden complementar cada fase.

FASE 1: ECAs de eficacia

OBJETIVO: Estudiar la eficacia de una intervención en condiciones ideales. Su objetivo es estudiar la relación causal entre la intervención y su efecto fisiológico. Se podrían incluir experimentos que no se realizan en humanos, estudios para identificar los posibles

efectos adversos de una intervención y para encontrar la dosis necesaria para conseguir un cambio fisiológico [136].

MÉTODOS:

- muestras muy homogéneas en sujetos sanos excluyendo sujetos que se intuye pueden no responder bien a la intervención y excluyendo co-morbididades. Muestra pequeña.
- Duración del estudio corta.
- Entorno con muchos recursos.
- La intervención se realiza en condiciones muy controladas y de forma muy protocolizada.
- Terapeutas con experiencia y especializados.
- La intervención control se realiza con placebo.
- Las variables de resultados suelen ser medidas fisiológicas, a muy corto plazo.
- enmascaramiento de los participantes y evaluadores externos.
- Reclutamiento de los participantes a demanda de las necesidades del estudio.
- No se permiten co-intervenciones.

VALIDEZ:

- Alta validez interna
- Poca validez externa

TIPOS DE ESTUDIO [6]:

- Mecanismos neurológicos
- Mecanismos pneuroinmunes
- Relación entre los mecanismos biomecánicos y neurofisiológicos
- Mecanismos neurovasculares
- Mecanismos Translacionales
- Factores contextuales en relación a distintos mecanismos
- Se podrían incluir experimentos que no se realizan en humanos

OTROS DISEÑOS POSIBLES EN ESTA FASE:

- ECAs no controlados con grupo control

FASE 2: ECAs híbridos de efectividad

Se realizarán ECAs híbridos de efectividad sobre un tema en concreto cuando existan ECAs de eficacia que prueben la eficacia de la intervención en condiciones ideales. Es a partir de este punto cuando se introducirán elementos pragmáticos al estudio.

OBJETIVO: Introducir elementos pragmáticos necesarios en intervenciones complejas como la TM especialmente referentes a la intervención. Ayudar a describir efectos adversos en patología. Replicar o verificar los efectos fisiológicos derivados de la intervención cuando se incluyen factores contextuales propios del encuentro paciente-terapeuta.

MÉTODOS:

- enmascaramiento de los participantes y evaluadores externos
- Intervención no protocolizada y centrada en las necesidades del paciente
- Intervención control con placebo
- Pueden ser muestras homogéneas en pacientes con patología, con criterios de exclusión rigurosos, pero especificando que la transferibilidad de los resultados será a ese tipo de población en concreto. No incluir sujetos sanos. Muestras pequeñas
- Entorno parecido a la práctica real
- Terapeutas con experiencia
- Reclutamiento de los participantes a través del mismo sistema que se reclutan pacientes en la práctica clínica
- Las variables pueden ser fisiológicas (para no interferir en el enmascaramiento de los asesores externos en el caso de PROMs). Valorar efectos a corto-medio plazo

VALIDEZ:

- Alta validez interna

- Aceptable validez externa

TIPOS DE ESTUDIO:

- ECAs intervención experimental vs placebo
- Se pueden replicar estudios de la FASE 1 añadiendo elementos pragmáticos en la intervención y su organización

OTROS DISEÑOS POSIBLES EN ESTA FASE:

- Cluster ECA
- ECAs lista de espera

FASE 3: Estudios de efectividad comparativa

OBJETIVOS: Comparar la efectividad de una intervención sobre otra y con placebo en condiciones de práctica habitual. El motivo de añadir un tercer grupo con control placebo se fundamenta en que el placebo se considera parte del efecto modulador. Además, sin un enmascaramiento adecuado los estudios con PROMs tienden a mostrar tamaños de efecto mayores [136,153]

Introduce más elementos pragmáticos a parte de la intervención. En esta fase ya se ha demostrado la eficacia y la seguridad de la intervención.

MÉTODOS:

- enmascaramiento de los participantes y evaluadores externos
- Intervención no protocolizada y centrada en las necesidades del paciente
- Intervención control con otra intervención establecida y otra con placebo
- Población heterogénea con pocos criterios de exclusión
- Muestra más grande que en las fases anteriores
- Realizar análisis estadísticos por subgrupos.
- Entorno parecido a la práctica real, considerar estudios multicéntricos.
- Terapeutas con experiencia (reportarlo en métodos y/o limitaciones)
- Reclutamiento de los participantes a través del mismo sistema que se reclutan

pacientes en la práctica clínica

- Variables relevantes para los participantes PROMs y valorar efectos a corto, medio y largo plazo
- Se puede añadir estudio de coste-efectividad

VALIDEZ:

- Aceptable validez interna
- Alta validez externa

TIPO DE ESTUDIOS:

- ECA de efectividad 3-arm design: control con otra intervención + placebo
- Estudios de coste-efectividad

OTROS DISEÑOS EN ESTA FASE:

- Considerar métodos alternativos a la aleatorización simple Estudios de preferencia, aleatorización por grupos (cluster randomisation), diseños stepped-wedge, enrichment methods
- Estudios de Cohortes
- Estudios Métodos mixtos, cuantitativos y cualitativos

FASE 4: Estudios pragmáticos

OBJETIVO: Estudiar el efecto de la intervención en situaciones reales. En esta fase se pretende implementar la intervención como opción terapéutica.

MÉTODOS:

- Open-label
- Intervención no protocolizada y centrada en las necesidades del paciente
- Intervención control usual care (la intervención que se acostumbra hacer para la

condición patológica en el sistema sanitario en concreto)

- Población heterogénea con criterios de exclusión propios de la patología de estudio
- Muestras relativamente grandes
- Entorno parecido a la práctica real
- Terapeutas con y sin experiencia
- Reclutamiento de los participantes a través del mismo sistema que se reclutan pacientes en la práctica clínica
- Variables relevantes para los participantes PROMs i variables objetivables para reducir el sesgo que supondría el enmascaramiento del participante que no existe y del evaluador, que en el caso de PROMs tampoco existiría.
- Valorar efectos a largo plazo

VALIDEZ:

- Poca validez interna (aunque probada en fases anteriores)
- Alta validez externa

TIPO DE ESTUDIOS:

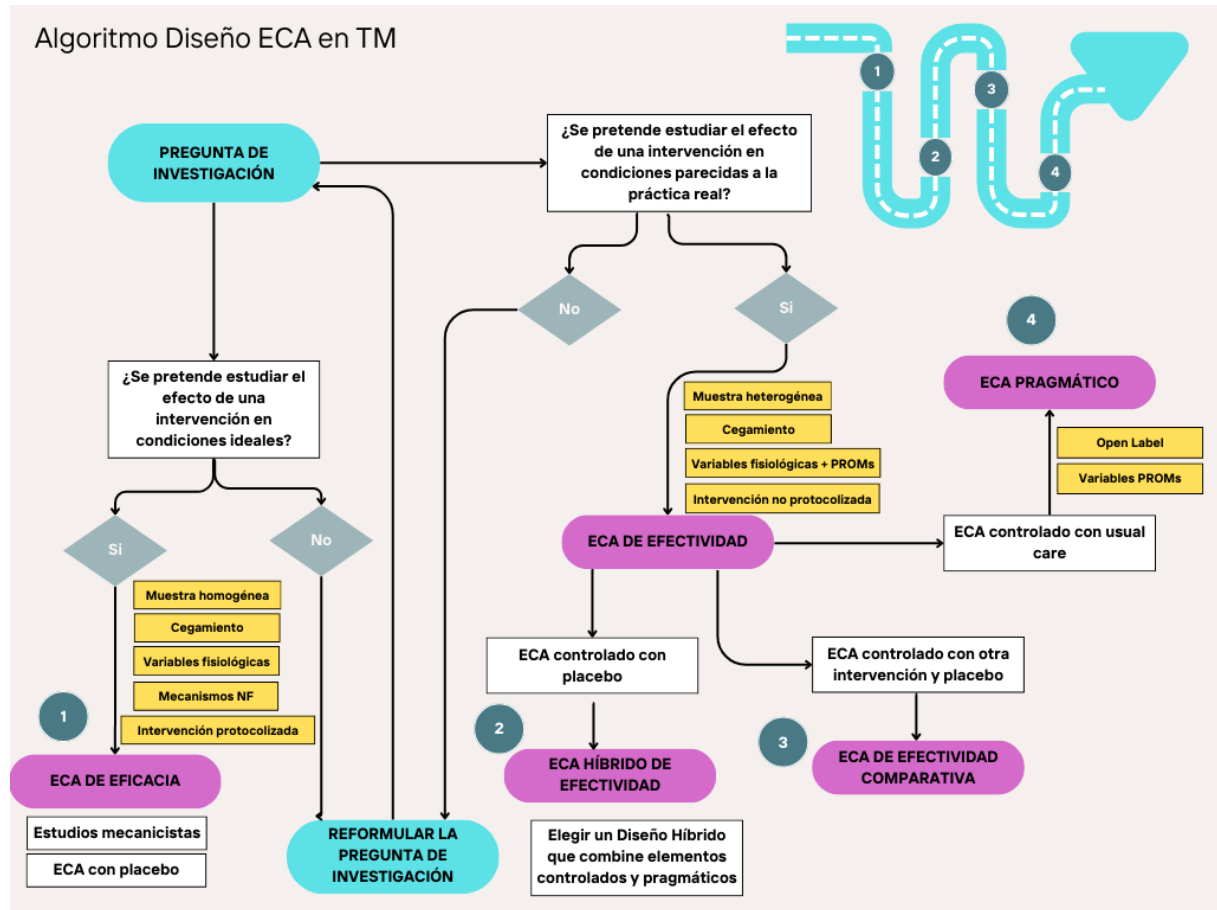
Estudios de efectividad auto-denominados pragmáticos

OTROS DISEÑOS POSIBLES EN ESTA FASE:

Estudios observacionales

Figura 9. Propuesta de algoritmo para desarrollar y evaluar las intervenciones de TM

Algoritmo Diseño ECA en TM



1: Fase 1, 2: Fase 2, 3: Fase 3, 4: Fase 5, ECA: ensayo controlado aleatorio, NF: neurofisiológicos, PROMs (patient reported outcomes)

PROPUESTA DE PROTOCOLO DELPHI

A framework and guidance to conduct and report high-quality randomised controlled trials on manual therapy interventions. A Delphi study

Roura SI, Álvarez G, Hohenschurz-Schmidt D, Sitjà M, Urrutia G

INTRODUCTION

Delphi studies aim to develop an expert-based judgement about an epistemic question. This is

based on the assumption that a group of experts will produce a more valid result than a judgement given by an individual expert ([Niederberger and Spranger 2020](#)). These techniques are frequently employed to identify consensus and integrate disparate bodies of knowledge about intricate matters where the available evidence is inconclusive and incomplete. This is achieved through an iterative process ([Linstone 1979](#)). Moreover, it is advised that standards or guidelines be established for theoretical and methodological issues ([Negrini et al. 2020](#); [Jorm 2015](#)). Without a consensus on the quality criteria ([Guzys et al. 2015](#)), this protocol was developed following the Guidance on Conducting and Reporting Delphi Studies (CREDES). ([Jünger et al. 2017](#)).

Manual therapy (MT) represents a therapeutic approach employed by a range of health professionals (physiotherapists, osteopaths, chiropractors, massage therapists), typically for treating diverse health issues, particularly musculoskeletal conditions ([Alvarez et al. 2020](#); [Oliveira et al. 2018](#); [Qaseem et al. 2017](#); [Bishop et al. 2015](#); [National Guideline Centre \(UK\) 2021](#); [Keter et al. 2024](#); [Draper-Rodi et al. 2024](#)). Manual therapy interventions encompass a range of techniques, including soft tissue techniques, joint mobilisations or manipulations, massage, light touch, myofascial release, nerve manipulation, tension/counterstrain and acupressure ([Alvarez et al. 2020](#); [Keter et al. 2024](#)). These techniques are regarded as complex interventions since they entail multiple factors that extend beyond the straightforward application of a single technique ([Wade, Smeets, and Verbunt 2010](#); [Bialosky et al. 2018](#); [Bialosky, Bishop, and Penza 2017](#)). Those multiple factors include use of touch as a diagnostic and treatment tool, patient-therapist interaction (placebo effect, contextual effects), patient-reported outcomes, and patient-centred interventions) ([Bishop et al. 2015](#); [Draper-Rodi et al. 2024](#); [Bialosky, Bishop, and Penza 2017](#); [Geri et al. 2019](#); [Ezzatvar et al. 2024](#); [Wade, Smeets, and Verbunt 2010](#); [Fennelly et al. 2018](#)). The complexity mentioned above can result in challenges in comprehending and assessing the underlying mechanism of action of the intervention, as well as determining its efficacy in actual practice ([Craig et al. 2008](#); [Dieppe 2004](#); [Ford and Norrie 2016](#); [Bennell et al. 2011](#); [Bialosky et al. 2018](#)). While different types of studies can be employed to address the complexity of some interventions ([Walach et al. 2006](#); [Minary et al. 2019](#)), RCTs remain the most appropriate design for objectifying the effect of an intervention. A search of the literature on musculoskeletal manipulations using the term 'musculoskeletal manipulations' in the field of TM, corresponding to November 2024, revealed 2709 publications in the last five years, of which 895 were randomised clinical trials (RCTs). A substantial body of literature on this topic has been subjected to analysis by different authors in various methodological reviews ([S. Roura et al. 2024](#); [Sonia Roura et al. 2021](#); [Alvarez et al. 2021a](#); [Núñez-Cortés et al. 2021](#); [Alvarez et al. 2020](#); [D. Hohenschurz-Schmidt, Draper-Rodi, Vase, Scott, McGregor, Soliman, MacMillan, Olivier, Cherian, Corcoran, Abbey, Freigang, Chan, Phalip, Sørensen, Delafin, Baptista, Medforth, Ruffini, Andresen, Ytier, Ali, Hobday, Santosa, et al. 2022](#); [D. Hohenschurz-Schmidt, Draper-Rodi, Vase, Scott, McGregor, Soliman, MacMillan, Olivier, Cherian, Corcoran, Abbey, Freigang, Chan, Phalip, Sørensen, Delafin, Baptista, Medforth, Ruffini, Andresen, Ytier, Ali, Hobday, Ngurah Agung Adhiyoga Santosa, et al. 2022](#); [D. Hohenschurz-Schmidt et al. 2021](#); [D. Hohenschurz-Schmidt, Phalip, et al. 2023](#); [Innocenti et al. 2022](#); [Daniel Maddox et al., n.d.](#); [Adams et al. 2024](#); [Gonzalez et al. 2018](#); [Nascimento et al. 2019](#); [Moseley et al. 2014](#)). These studies have reached a consensus that, in general, TM studies tend to be of low quality, that the reporting of the studies is insufficient ([Núñez-Cortés et al. 2021](#); [Alvarez et](#)

al. 2020; Innocenti et al. 2022), that the applicability of the research to real clinical practice is doubtful (Daniel Maddox et al., n.d.; D. Hohenschurz-Schmidt et al. 2021; S. Roura et al. 2024; Adams et al. 2024) and that the use of the explanatory and effectiveness RCTs designs is often misunderstood. This ultimately leads to several scenarios, including clinicians' low utilisation of evidence and the lack of conclusive findings in multiple systematic reviews. In light of the above considerations, questions still need to be answered about designing and reporting a high-quality trial in MT and placing it in the explanatory-pragmatic continuum regarding a precise research question. Furthermore, guidelines to help researchers are lacking in the field of MT. As done in other CIs (Zhang et al. 2022), an international Delphi panel of manual therapists, researchers in the MT field and researchers in methodology will aim to develop guidance on conducting high-quality research in MT. A discussion will be framed with essential features to consider when designing trials.

Study purposes

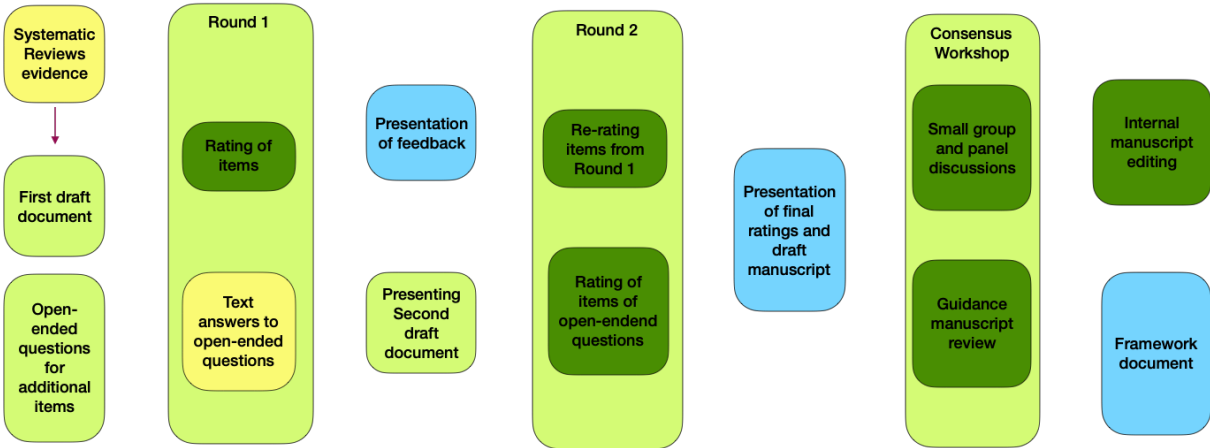
The primary aim of this Delphi process is to establish important considerations for the design and reporting of RCTs in MT according to the research question and to propose a framework to help MT researchers.

METHODS

As literature shows several shortcomings and misunderstandings in MT research, a multi-round international modified Delphi study, informed by several systematic reviews (S. Roura et al. 2024; Alvarez, Cerritelli, and Urrutia 2016; Alvarez et al. 2021b; Núñez-Cortés et al. 2021; Alvarez et al. 2020; Wartolowska et al. 2023; D. Hohenschurz-Schmidt, Vase, et al. 2023; D. J. Hohenschurz-Schmidt et al. 2023; D. Hohenschurz-Schmidt, Draper-Rodi, Vase, Scott, McGregor, Soliman, MacMillan, Olivier, Cherian, Corcoran, Abbey, Freigang, Chan, Phalip, Sørensen, Delafin, Baptista, Medforth, Ruffini, Andresen, Ytier, Ali, Hobday, Santosa, et al. 2022; D. Hohenschurz-Schmidt, Draper-Rodi, Vase, Scott, McGregor, Soliman, MacMillan, Olivier, Cherian, Corcoran, Abbey, Freigang, Chan, Phalip, Sørensen, Delafin, Baptista, Medforth, Ruffini, Andresen, Ytier, Ali, Hobday, Ngurah Agung Adhiyoga Santosa, et al. 2022; D. Hohenschurz-Schmidt et al. 2021; Innocenti et al. 2022; Adams et al. 2024; Daniel Maddox et al., n.d.; Cook and Thigpen 2019; Keter et al. 2024; Ezzatvar et al. 2024; Rossettini et al. 2020), will be conducted aiming to achieve consensus on how to design and report high-quality clinical trials in MT.

The Delphi technique will be adapted to allow for eliciting panellists’ opinions and considering the results from several systematic reviews of methods. The multi-round Delphi process is shown in Figure 1. Panellists will be presented with a first draft document created by the steering committee and informed by the results of the cited systematic reviews. The first round will consist of a survey and open-ended questions. Following the initial round of ratings and comments, the draft document will be refined in light of the expert panel's input. A new document will be drafted and distributed to all participants in a second round. After round 2, a virtual consensus meeting will be scheduled for the finalisation of a recommendation framework, as has been done by other research groups with comparable objectives (Evers et al. 2018).

Figure 1. Flow diagram illustrating the consensus-finding and guidance-development process. Yellow elements represent inputs into the consensus process. Light green elements are part of the process, with darker green signifying rating and evaluation procedures. Blue are interim or final outputs. Methods for the systematic review and the workshop are presented elsewhere



Protocol registration

It will be registered in the Open Science Framework before recruitment.

Development of the Delphi questionnaire

After two online workshops, a steering committee will develop a First draft Framework (FdF) based on pertinent literature about research methodology in CI, especially in MT (see references above). For the panellists to work on the FdF, the steering committee will develop a Delphi questionnaire to rate each item of the FdF. Moreover,

The FdF will contain information regarding:

1. Design items
 - a. Research question
 - b. Rational and intention of the trial
 - c. Eligibility of participants
 - d. Recruitment
 - e. Setting
 - f. Experimental intervention
 - g. Control intervention
 - h. Therapists
 - i. Follow-up and adherence
 - j. Outcome measures
 - k. Statistical analysis

2. Reporting items

The survey items can be grouped as follows, and the entire survey instrument can be found in Appendix 1:

1. Rating of each item of the FdF: A 9-point Likert scale will be used, anchored at Strongly Disagree (1) and Strongly Agree (9). Panellists will be asked to rate each item, answering the following question followed by a single survey item: Should researchers planning, conducting or reporting research in manual therapy consider the following item? (Table 2)

Table 2: The original RAND UCLA rating scale was developed for Delphi forecasting methods and employing a 9-point Likert instrument (Rescher 1998).

Strongly Disagree	Disagree	Moderately Disagree	Mildly Disagree	Undecided	Mildly Agree	Moderately Agree	Agree	Strongly Agree
1	2	3	4	5	6	7	8	9

2, Open-ended questions

a. Open-ended questions of round 1

As per recommendation, the first round will include a brief set of open-ended questions after each item of the FdF. These questions allow the panellist to add some considerations not collected in the survey. They read as follows (still to develop):

What makes the difference between a pragmatic and an effectiveness trial in an MT trial where development phases are not considered per pharmacological research?

How can trialists address possible differences in treatment effects between groups?

Are the results of a unicentric study informative?

Which comparator groups are suitable for the different trial designs? What kind of outcomes?

What is an appropriate follow-up?

How important is blinding?

How should trialists deal with participants' adherence?

The survey and the open-ended questions will be piloted on a small sample of participants.

Feedback and information between rounds

The steering committee will collate all the scores and comments obtained in the initial phase and prepare a Second draft Framework (SdF) to commence the second phase of the process.

Definition of consensus

The Delphi survey will end after two rounds, with individual items removed from the rating process once 80% of the entire sample rated them as not worth considering (1-3) or considering (7-9). 80% agreement is a standard criterion for consensus in Delphi techniques ([Jünger et al. 2017](#)). This consensus definition will inform the strength of recommendations for each item in the final guidance document and the extent to which items will be discussed at the live workshop.

Consensus workshop

As the aim is to produce practicable guidance on the methodological design and reporting of MT trials, the results of the Delphi exercise will be discussed in a consensus workshop attended by Delphi participants. At this workshop, which will take place 4-8 weeks after the final Delphi round, the workshop participants will present and discuss a manuscript outlining the recommendations based on Delphi results. Procedures at the workshop are aligned with processes adopted by the James Lind Alliance for identifying health research priorities (<https://www.jla.nihr.ac.uk/jla-guidebook/chapter-8/final-priority-setting.htm>). Before the meeting, the proposed manuscript will be circulated, and participants will be invited to note any amendments they want to propose at the meeting. They will be informed that the workshop focuses on gathering additional nuance on individual recommendations to make context-specific or otherwise qualified recommendations.

At the meeting, small-group discussions (max. 10 participants, equal proportions of trial researchers and clinicians) will produce a list of amendments, which the plenum will then discuss. Afterwards, anonymously, the same rating scale as during the Delphi exercise will be used to rate the agreement for each amendment. This workshop will be facilitated by steering group members who will introduce the ground rules for the meeting and moderate small-group and panel discussions to ensure that all participants are engaged and a few individuals do not dominate discussions.

The reformulated manuscript will be circulated to all Delphi attendants, who will be invited to make editorial changes and co-author it if they wish.

The consensus workshop will be held online, using the Zoom videoconferencing platform and at a convenient time for most participants. Accepting that there will be time-zone-dependent

difficulties, more than one workshop may have to be scheduled so that every panellist who wishes to attend can do so. The research committee will ensure that no workshop discussion dominates the manuscript development. Meetings will be audio-recorded to allow for an accurate representation of participants' views.

Selection of participants

Delphi panellists will be included in this study if they consent and fulfil the eligibility criteria below (Table 3). Broadly, we are looking for a sample composed of researchers and clinicians. The researchers' sample will include researchers with experience in MT and clinical trial methods.

Table 3. Eligibility criteria for panellists, recruiting into either of two expert groups.

Inclusion criteria	
Clinical trial researchers	<p>Either:</p> <ul style="list-style-type: none"> At least one peer-reviewed publication related to MT clinical trials (irrespective of author position and including systematic reviews or methods papers concerned with such trials) <p>Or:</p> <ul style="list-style-type: none"> Academic or professional training related to clinical trial methods, i.e., at least an undergraduate module at a registered teaching institution <i>and</i> professional background in MT, i.e., at least an undergraduate degree from a registered institution <i>or</i> research interest in clinical trials of MT
Methodological researchers	At least one peer-reviewed publication related to methodological aspects of non-pharmacological interventions
Exclusion criteria	
Inability to understand and write English	
Inability or unwillingness to complete up to two online surveys within the expected timeframe	

Sample size of Delphi panel

Formal sample size calculations are uncommon in Delphi designs, and sample sizes are commonly thought to lie between 10 and 20 experts (Hasson, Keeney, and McKenna 2000; McPherson, Reese, and Wendler 2018). Because it has been pointed out that the success of a Delphi technique depends on both adequate panel size and the qualification of panellists (McPherson, Reese, and Wendler 2018), the success or failure to recruit leading experts in the field may influence the decision to stop recruitment. Given that two distinct groups are recruited and the accepted minimum number of panellists is ten, at least 20 individuals must be recruited, and more if leading experts decline to participate.

Sampling strategy and recruitment rate

Purposive sampling will be employed, approaching corresponding authors of eligible publications, i.e., RCTs in MT and related methodological articles, using a standardised email invitation. Snowball sampling will be employed by inviting responders to nominate potentially eligible candidates further. To obtain a representative sample, corresponding authors of all trials identified by a systematic literature search will be contacted.

Deliberate attempts will be made to recruit a diverse sample regarding geographic location, gender, and career stage. No direct incentives for participation will be offered. However, participants can decide if they wish to be listed as contributors to the guidance document, which may act as an incentive.

Invitations and reminders

Invitations to participate in the Delphi will be emailed to pre-identified professional email addresses or individuals registering their interest with the study team. Recruitment emails will contain lay-language information about the study's nature and purposes and eligibility criteria for participation. Before starting recruitment, the emails will undergo an institutional ethics review. For ethical reasons and to enhance retention, potential participants will be informed of the estimated time required and their need to stay engaged (McPherson, Reese, and Wendler 2018).

A reminder will be sent to non-respondents 7 days after the initial email. Upon registering interest, the participant information sheet (PIS) and the consent form will be provided, and reminders will be sent if not returned within 7 days.

Participant retention and response rate to survey rounds

Next to insufficient or biased recruitment, the loss of survey panellists throughout the study is the most significant threat to the internal validity of a Delphi technique (McPherson, Reese, and Wendler 2018). Possible reasons for panellist attrition are survey workload, disinterest, discontent with the direction of the study or the management through the researcher, and external reasons (Sherwood Peter, Deery Margaret, and Jago Leo, n.d.). To ensure the trustworthiness of findings, retention of panellists should be high, at least 70%, per recommendations (REF). To increase response rates and improve retention, up to two reminder emails will be sent to non-responding panellists, deadlines for round completions will be extended if needed, and individualised thank-you messages will be sent upon round completion. Some authors note that response rates drop with increasing iterations, so this Delphi will be limited to a maximum of two rounds even if not all items reach consensus (McPherson, Reese, and Wendler 2018).

Administration of the survey and data management

Survey data will be collected and managed using XX??? electronic data capture tools hosted. REDCap (Research Electronic Data Capture) is a secure, web-based software platform designed to support data capture for research studies, providing an intuitive interface for validated data capture, audit trails for tracking data manipulation and export procedures, and automated export procedures for seamless data downloads to common statistical packages (Harris et al., 2009, 2019). Surveys will include identifiable information in the form of the participants' names so that during subsequent survey rounds each participant can be reminded of their ratings from the previous round (provided individually). However, data will be made de-identifiable for analysis purposes, and feedback provided to the entire group will be anonymous between rounds.

All data, including signed consent forms and participants' personal information, will be encrypted, password-protected, and stored within the Imperial College IT infrastructure. Specifically, data will be stored using Microsoft OneDrive and encrypted using nCrypted Cloud, as per Imperial college recommendation. Prior to commencing the project, a Data Privacy Impact Assessment (DPIA) will be conducted by an Imperial College data protection advisor, ensuring all processes be compliant with GDPR / DPA 2018.

Informed consent

Panellists will participate in the study as researchers based on their voluntary and informed decision. Upon declaring interest in the study and confirming eligibility, a participant information sheet (PIS) will be provided. The PIS contains all relevant information and will undergo institutional ethics review.

The informed consent process is based on the information provided in the PIS. Participants can discuss any remaining questions with a research team member before giving consent, either by

phone or email. The PIS is written in understandable English and has been tested with non-native speakers who can converse in English confidently. As being able to speak and read English is one of the eligibility criteria, no additional measures will be taken to make the consent process accessible to non-English speakers.

Participants will be free to withdraw from the study at any point by informing a research team member without giving a reason. Email addresses to do so will be provided. As part of the PIS, participants will be informed that withdrawing after one or more Delphi rounds will mean that their demographic and other data provided until that point will remain part of the research data, albeit non-identifiable. Retrospectively removing all data will not be possible because data from previous rounds will already have informed subsequent rounds. This is a common procedure in Delphi surveys.

Analysis

The results of this Delphi study will provide insights regarding the supposed value of specific elements of manual therapy interventions, as well as their design and reporting, to conduct high-quality research in this field. Readers will understand which of these items were not considered worthy of recommendation by a group of clinical trial researchers in MT and researchers in non-pharmacological trial methodology, as well as by both groups. Discussion of the result will also consider items for which rating disagreements were present throughout the survey, possibly highlighting areas for increased research needs.

Data for each item will be presented as follows. Supplementary material will report results for each round separately:

- Reached consensus for recommendation (as defined above)
 - Percentage of panellists per group and across groups rating 7-9 in the item's final round
- Reached consensus against recommendation (as defined above)
 - Percentage of panellists per group and across groups rating 1-3 in the item's final round
- Average rating and variance
- Whether an item's average score increased or decreased by one or more points after panellists were presented with the systematic review, or if it remained unchanged.

If indicated, recommendation items will be presented in a table or graph and grouped as above (see the beginning of methods).

Reporting and quality guidelines

This protocol was developed in line with the recommendations for the Conducting and REporting of Delphi Studies (CREDES), and the study will be reported accordingly ([Jünger et al. 2017](#)).

Social scientists are often aware of threats to the external validity of surveys, mainly due to sampling errors and response bias, leading to a non-representative sample. According to R. Bautista (2012), less attention is given to the threats to the internal validity of surveys, which he argues is essential ([Bautista 2012](#)). Total Survey Error (TSE) addresses both aspects by

highlighting sampling, coverage, nonresponse, and measurement errors ([Biemer 2010](#)). The present protocol seeks to minimise these errors by outlining appropriate methods.

Dissemination

The research output will be a guidance document for designing and reporting high-quality research in manual therapy. This document will outline the methods of guidance development as per above, present the findings from the Delphi technique, and discuss the adaptations made at the workshop. Final recommendations will be presented in an accessible table format, with descriptions and explanations for each recommended item. This manuscript will be submitted for peer review and publication in a scientific journal.

Ethics

Ethical considerations include the risk of coercing people into participation, which will be addressed by highlighting the voluntary nature of the involvement in the PIS. The complete anonymity of individual panellists can be ensured if desired. This includes anonymous participation in the online live workshop by allowing participants to enter the workshop with a pseudonym (pre-agreed with the study team) and without video. At the workshop, ethical concerns are professional and academic hierarchies that may appear intimidating to some and differences in how confident participants may feel in speaking publicly and in English. For these reasons, methods from the James Lind Alliance have been adopted, where small groups and deliberate moderation by trained research staff seek to minimise these risks and allow equal involvement of all participants.

No physical or psychological harm is expected, neither for the participants nor for the researchers, as no dangerous procedures are involved, and no personal or sensitive issues are expected to be discussed.

LIST OF ABBREVIATIONS

CI	Complex interventions
MT	Manual therapy
RCT	Randomised clinical trial
LBP	Low back pain
FdF	First Draft Framework
DsF	Second draft Framework
PIS	participant information sheet

Adams, Kyle R., Ayodeji O. Famuyide, Jodi L. Young, C. Daniel Maddox, and Daniel I. Rhon. 2024. "Pragmatism in Manual Therapy Trials for Knee Osteoarthritis: A Systematic Review." *Archives of Physiotherapy* 14 (February):1–10.

Alvarez, Gerard, Francesco Cerritelli, and Gerard Urrutia. 2016. "Using the Template for Intervention Description and Replication (TIDieR) as a Tool for Improving the Design and Reporting of Manual Therapy Interventions." *Manual Therapy* 24 (August):85–89.

Alvarez, Gerard, Rodrigo Núñez-Cortés, Ivan Solà, Mercè Sitjà-Rabert, Azahara Fort-Vanmeerhaeghe, Carles Fernández, Xavier Bonfill, and Gerard Urrútia. 2021a. "Sample Size, Study Length, and Inadequate Controls Were the Most Common Self-Acknowledged Limitations in Manual Therapy Trials: A Methodological Review." *Journal of Clinical Epidemiology* 130 (February):96–106.

— — —. 2021b. "Sample Size, Study Length, and Inadequate Controls Were the Most Common Self-Acknowledged Limitations in Manual Therapy Trials: A Methodological Review." *Journal of Clinical Epidemiology* 130 (February):96–106.

Alvarez, Gerard, Ivan Solà, Mercè Sitjà-Rabert, Azahara Fort-Vanmeerhaeghe, Ignasi Gich, Carles Fernández, Xavier Bonfill, and Gerard Urrútia. 2020. "A Methodological Review Revealed That Reporting of Trials in Manual Therapy Has Not Improved over Time." *Journal of Clinical Epidemiology* 121 (May):32–44.

Bautista, René. 2012. "An Overlooked Approach in Survey Research: Total Survey Error." In *Handbook of Survey Methodology for the Social Sciences*, edited by Lior Gideon, 37–49. New York, NY: Springer New York.

Bennell, Kim L., Thorlene Egerton, Yong-Hao Pua, J. Haxby Abbott, Kevin Sims, and Rachelle Buchbinder. 2011. "Building the Rationale and Structure for a Complex Physical Therapy Intervention within the Context of a Clinical Trial: A Multimodal Individualized Treatment for Patients with Hip Osteoarthritis." *Physical Therapy* 91 (10): 1525–41.

Bialosky, Joel E., Jason M. Beneciuk, Mark D. Bishop, Rogelio A. Coronado, Charles W. Penza, Corey B. Simon, and Steven Z. George. 2018. "Unraveling the Mechanisms of Manual Therapy: Modeling an Approach." *The Journal of Orthopaedic and Sports Physical Therapy* 48 (1): 8–18.

Bialosky, Joel E., Mark D. Bishop, and Charles W. Penza. 2017. "Placebo Mechanisms of Manual Therapy: A Sheep in Wolf's Clothing?" *The Journal of Orthopaedic and Sports Physical Therapy* 47 (5): 301–4.

Biemer, Paul P. 2010. "Total Survey Error: Design, Implementation, and Evaluation." *Public Opinion Quarterly* 74 (5): 817–48.

Bishop, Mark D., Rafael Torres-Cueco, Charles W. Gay, Enrique Lluch-Girbés, Jason M. Beneciuk, and Joel E. Bialosky. 2015. "What Effect Can Manual Therapy Have on a Patient's Pain Experience?" *Pain Management* 5 (6): 455–64.

Cook, Chad E., and Charles A. Thigpen. 2019. "Five Good Reasons to Be Disappointed with Randomized Trials." *The Journal of Manual & Manipulative Therapy* 27 (2): 63–65.

Craig, Peter, Paul Dieppe, Sally Macintyre, Susan Michie, Irwin Nazareth, Mark Petticrew, and Medical Research Council Guidance. 2008. "Developing and Evaluating Complex Interventions: The New Medical Research Council Guidance." *BMJ* 337 (September):a1655.

Daniel Maddox, C., Joshua A. Subialka, Jodi L. Young, and Daniel I. Rhon. n.d. "TITLE: Over Half of Clinical Trials of Mobilization and Manipulation for Patients with Low Back

Pain May Have Limited Real-World Applicability. A Systematic Review Pf 132 Clinical Trials."

Dieppe, Paul. 2004. "Complex Interventions." Musculoskeletal Care 2 (3): 180–86.

Draper-Rodi, Jerry, Dave Newell, Mary F. Barbe, and Joel Bialosky. 2024. "Integrated Manual Therapies: IASP Taskforce Viewpoint." Pain Reports (Baltimore, Md.) 9 (6): e1192.

Evers, Andrea W. M., Luana Colloca, Charlotte Blease, Marco Annoni, Lauren Y. Atlas, Fabrizio Benedetti, Ulrike Bingel, et al. 2018. "Implications of Placebo and Nocebo Effects for Clinical Practice: Expert Consensus." Psychotherapy and Psychosomatics. <https://doi.org/10.1159/000490354>.

Ezzatvar, Yasmin, Lirios Dueñas, Mercè Balasch-Bernat, Enrique Lluch-Girbés, and Giacomo Rossetini. 2024. "Which Portion of Physiotherapy Treatments' Effect Is Attributable to Contextual Effects in People with Musculoskeletal Pain?: A Meta-Analysis of Randomised Placebo-Controlled Trials." The Journal of Orthopaedic and Sports Physical Therapy, April, 1–28.

Fennelly, Orna, Catherine Blake, François Desmeules, Diarmuid Stokes, and Caitriona Cunningham. 2018. "Patient-Reported Outcome Measures in Advanced Musculoskeletal Physiotherapy Practice: A Systematic Review." Musculoskeletal Care 16 (1): 188–208.

Ford, Ian, and John Norrie. 2016. "Pragmatic Trials." New England Journal of Medicine. <https://doi.org/10.1056/nejmra1510059>.

Geri, Tommaso, Antonello Viceconti, Marco Minacci, Marco Testa, and Giacomo Rossetini. 2019. "Manual Therapy: Exploiting the Role of Human Touch." Musculoskeletal Science and Practice 44 (December):102044.

Gonzalez, Gabrielle Zoldan, Anne M. Moseley, Christopher G. Maher, Dafne Port Nascimento, Lucíola da Cunha Menezes Costa, and Leonardo Oliveira Costa. 2018. "Methodologic Quality and Statistical Reporting of Physical Therapy Randomized Controlled Trials Relevant to Musculoskeletal Conditions." Archives of Physical Medicine and Rehabilitation 99 (1): 129–36.

Guzys, Diana, Virginia Dickson-Swift, Amanda Kenny, and Guinever Threlkeld. 2015. "Gadamerian Philosophical Hermeneutics as a Useful Methodological Framework for the Delphi Technique." International Journal of Qualitative Studies on Health and Well-Being. <https://doi.org/10.3402/qhw.v10.26291>.

Hasson, F., S. Keeney, and H. McKenna. 2000. "Research Guidelines for the Delphi Survey Technique." Journal of Advanced Nursing 32 (4): 1008–15.

Hohenschurz-Schmidt, David, Dr Jerry Draper-Rodi, Prof Lene Vase, Dr Whitney Scott, Prof Alison McGregor, Dr Nadia Soliman, Andrew MacMillan, Axel Olivier, Cybill Ann Cherian, Daniel Corcoran, Dr Hilary Abbey, Sascha Freigang, Jessica Chan, Jules Phalip, Lea Nørgaard Sørensen, Maite Delafin, Margarida Baptista, Naomi R. Medforth, Nuria Ruffini, Stephanie Skøtt Andresen, Sylvain Ytier, Dorota Ali, Harriet Hobday, Anak Agung Ngurah Agung Adhiyoga Santosa, et al. 2022. "Blinding and Sham Control Methods in Trials of Physical, Psychological, and Self-Management Interventions for Pain (article II): A Meta-Analysis Relating Methods to Trial Results." Pain, May, 10.1097/j.pain.0000000000002730.

Hohenschurz-Schmidt, David, Jerry Draper-Rodi, Lene Vase, Whitney Scott, Alison McGregor, Nadia Soliman, Andrew MacMillan, Axel Olivier, Cybill Ann Cherian, Daniel Corcoran, Hilary Abbey, Sascha Freigang, Jessica Chan, Jules Phalip, Lea Nørgaard

Sørensen, Maite Delafin, Margarida Baptista, Naomi R. Medforth, Nuria Ruffini, Stephanie Skøtt Andresen, Sylvain Ytier, Dorota Ali, Harriet Hobday, Anak Agung Ngurah Agung Santosa, et al. 2022. "Blinding and Sham Control Methods in Trials of Physical, Psychological, and Self-Management Interventions for Pain (Article I): A Systematic Review and Description of Methods." *Pain*. <https://doi.org/10.1097/j.pain.0000000000002723>.

Hohenschurz-Schmidt, David J., Dan Cherkin, Andrew S. C. Rice, Robert H. Dworkin, Dennis C. Turk, Michael P. McDermott, Matthew J. Bair, et al. 2023. "Research Objectives and General Considerations for Pragmatic Clinical Trials of Pain Treatments: IMMPACT Statement." *Pain*, March. <https://doi.org/10.1097/j.pain.0000000000002888>.

Hohenschurz-Schmidt, David, Bethea A. Kleykamp, Jerry Draper-Rodi, Jan Vollert, Jessica Chan, Mckenzie Ferguson, Ewan McNicol, et al. 2021. "Pragmatic Trials of Pain Therapies: A Systematic Review of Methods." *Pain*. https://journals.lww.com/pain/abstract/9000/pragmatic_trials_of_pain_therapies_a_systematic.98036.aspx.

Hohenschurz-Schmidt, David, Jules Phalip, Jessica Chan, Greta Gauhe, Nadia Soliman, Jan Vollert, Sigrid Juhl Lunde, and Lene Vase. 2023. "Placebo Analgesia in Physical and Psychological Interventions: Systematic Review and Meta-Analysis of Three-Armed Trials." *European Journal of Pain*, November. <https://doi.org/10.1002/ejp.2205>.

Hohenschurz-Schmidt, David, Lene Vase, Whitney Scott, Marco Annoni, Oluwafemi K. Ajayi, Jürgen Barth, Kim Bennell, et al. 2023. "Recommendations for the Development, Implementation, and Reporting of Control Interventions in Efficacy and Mechanistic Trials of Physical, Psychological, and Self-Management Therapies: The CoPPS Statement." *BMJ* 381 (May):e072108.

Innocenti, Tiziano, Silvia Giagio, Stefano Salvioli, Daniel Feller, Silvia Minnucci, Fabrizio Brindisino, Wilhelmina IJzelenberg, Raymond Ostelo, and Alessandro Chiarotto. 2022. "Completeness of Reporting Is Suboptimal in Randomized Controlled Trials Published in Rehabilitation Journals, With Trials With Low Risk of Bias Displaying Better Reporting: A Meta-Research Study." *Archives of Physical Medicine and Rehabilitation*. <https://doi.org/10.1016/j.apmr.2022.01.156>.

Jorm, Anthony F. 2015. "Using the Delphi Expert Consensus Method in Mental Health Research." *The Australian and New Zealand Journal of Psychiatry* 49 (10): 887–97.

Jünger, Saskia, Sheila A. Payne, Jenny Brine, Lukas Radbruch, and Sarah G. Brearley. 2017. "Guidance on Conducting and REporting DELphi Studies (CREDES) in Palliative Care: Recommendations Based on a Methodological Systematic Review." *Palliative Medicine* 31 (8): 684–706.

Keter, Damian L., Jennifer A. Bent, Joel E. Bialosky, Carol A. Courtney, Jorge E. Esteves, Martha Funabashi, Samuel J. Howarth, et al. 2024. "An International Consensus on Gaps in Mechanisms of Forced-Based Manipulation Research: Findings from a Nominal Group Technique." *The Journal of Manual & Manipulative Therapy* 32 (1): 111–17.

Linstone, Harold A. 1979. *The Delphi Method: Techniques and Applications*. Reading. Adison Wesley.

McPherson, Sara, Cynthia Reese, and M. Cecilia Wendler. 2018. "Methodology Update: Delphi Studies." *Nursing Research* 67 (5): 404–10.

Minary, Laetitia, Justine Trompette, Joëlle Kivits, Linda Cambon, Cyril Tarquinio, and

François Alla. 2019. "Which Design to Evaluate Complex Interventions? Toward a Methodological Framework through a Systematic Review." *BMC Medical Research Methodology* 19 (1): 92.

Moseley, Anne M., Mark R. Elkins, Lee Janer-Duncan, and Julia M. Hush. 2014. "The Quality of Reports of Randomized Controlled Trials Varies between Subdisciplines of Physiotherapy." *Physiotherapy Canada. Physiotherapie Canada* 66 (1): 36–43.

Nascimento, Dafne Port, Gabrielle Zoldan Gonzalez, Amanda Costa Araujo, Anne M. Moseley, Chris G. Maher, and Leonardo Oliveira Pena Costa. 2019. "Eight in Every 10 Abstracts of Low Back Pain Systematic Reviews Presented Spin and Inconsistencies With the Full Text: An Analysis of 66 Systematic Reviews." *The Journal of Orthopaedic and Sports Physical Therapy* 50 (1): 17–23.

National Guideline Centre (UK). 2021. Evidence Review for Manual Therapy for Chronic Primary Pain. National Institute for Health and Care Excellence (NICE).

Negrini, Stefano, Susan Armijo-Olivo, Michele Patrini, Walter R. Frontera, Allen W. Heinemann, Wendy Machalicek, John Whyte, Chiara Arienti, and RCTrack Promoters. 2020. "The Randomized Controlled Trials Rehabilitation Checklist: Methodology of Development of a Reporting Guideline Specific to Rehabilitation." *American Journal of Physical Medicine & Rehabilitation / Association of Academic Physiatrists* 99 (3): 210–15.

Niederberger, Marlen, and Julia Spranger. 2020. "Delphi Technique in Health Sciences: A Map." *Frontiers in Public Health* 8 (September):457.

Núñez-Cortés, Rodrigo, Gerard Alvarez, Javier Pérez-Bracchiglione, Rosa Cabanas-Valdés, Jordi Calvo-Sanz, Xavier Bonfill, and Gerard Urrutia. 2021. "Reporting Results in Manual Therapy Clinical Trials: A Need for Improvement." *International Journal of Osteopathic Medicine: IJOM* 0 (0). <https://doi.org/10.1016/j.ijosm.2021.06.002>.

Oliveira, Crystian B., Chris G. Maher, Rafael Z. Pinto, Adrian C. Traeger, Chung-Wei Christine Lin, Jean-François Chenot, Maurits van Tulder, and Bart W. Koes. 2018. "Clinical Practice Guidelines for the Management of Non-Specific Low Back Pain in Primary Care: An Updated Overview." *European Spine Journal: Official Publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society* 27 (11): 2791–2803.

Qaseem, Amir, Timothy J. Wilt, Robert M. McLean, Mary Ann Forciea, and for the Clinical Guidelines Committee of the American College of Physicians. 2017. "Noninvasive Treatments for Acute, Subacute, and Chronic Low Back Pain: A Clinical Practice Guideline From the American College of Physicians." *Annals of Internal Medicine*. <https://doi.org/10.7326/m16-2367>.

Rescher, Nicholas. 1998. *Predicting the Future: An Introduction to the Theory of Forecasting*. SUNY Press.

Rossettini, Giacomo, Eleonora Maria Camerone, Elisa Carlino, Fabrizio Benedetti, and Marco Testa. 2020. "Context Matters: The Psychoneurobiological Determinants of Placebo, Nocebo and Context-Related Effects in Physiotherapy." *Archives of Physiotherapy* 10 (June):11.

Roura, S., G. Alvarez, D. Hohenschurz-Schmidt, I. Solà, R. Núñez-Cortés, J. Bracchiglione, C. Fernández-Jané, et al. 2024. "Lack of Pragmatic Attitude of Self-Labelled Pragmatic Trials on Manual Therapy: A Methodological Review." *BMC Medical Research Methodology* 24 (1). <https://doi.org/10.1186/s12874-024-02393-1>.

Roura, Sonia, Gerard Álvarez, Ivan Solà, and Francesco Cerritelli. 2021. "Do Manual Therapies Have a Specific Autonomic Effect? An Overview of Systematic Reviews." PloS One 16 (12): e0260642.

Sherwood Peter, Deery Margaret, and Jago Leo. n.d. "Not Another Survey! Exploring the Issues in Conducting a Webbased Delphi Survey." CAUTHE 2006: To the City and Beyond, 1302–15.

Wade, Derick T., Rob J. E. M. Smeets, and Jeanine A. Verbunt. 2010. "Research in Rehabilitation Medicine: Methodological Challenges." Journal of Clinical Epidemiology 63 (7): 699–704.

Walach, Harald, Torkel Falkenberg, Vinjar Fønnebø, George Lewith, and Wayne B. Jonas. 2006. "Circular instead of Hierarchical: Methodological Principles for the Evaluation of Complex Interventions." BMC Medical Research Methodology 6 (June):29.

Wartolowska, Karolina A., David Hohenschurz-Schmidt, Lene Vase, and Jeffrey K. Aronson. 2023. "The Importance of Using Placebo Controls in Nonpharmacological Randomised Trials." Pain 164 (5): 921–25.

Zhang, Yu-Qing, Rui-Min Jiao, Claudia M. Witt, Lixing Lao, Jian-Ping Liu, Lehana Thabane, Karen J. Sherman, et al. 2022. "How to Design High Quality Acupuncture Trials—a Consensus Informed by Evidence." BMJ 376 (March). <https://doi.org/10.1136/bmj-2021-067476>.

9. BIBLIOGRAFÍA

- [1] Álvarez G, Solà I, Sitjà-Rabert M, Fort-Vanmeerhaeghe A, Gich I, Fernández C, et al. A methodological review revealed that reporting of trials in manual therapy has not improved over time. *J Clin Epidemiol* 2020;121:32–44.
- [2] Oliveira CB, Maher CG, Pinto RZ, Traeger AC, Lin C-WC, Chenot J-F, et al. Clinical practice guidelines for the management of non-specific low back pain in primary care: an updated overview. *Eur Spine J* 2018;27:2791–803.
- [3] Qaseem A, Wilt TJ, McLean RM, Forciea MA, for the Clinical Guidelines Committee of the American College of Physicians. Noninvasive Treatments for Acute, Subacute, and Chronic Low Back Pain: A Clinical Practice Guideline From the American College of Physicians. *Annals of Internal Medicine* 2017;166:514. <https://doi.org/10.7326/m16-2367>.
- [4] Bishop MD, Torres-Cueco R, Gay CW, Lluch-Girbés E, Beneciuk JM, Bialosky JE. What effect can manual therapy have on a patient's pain experience? *Pain Manag* 2015;5:455–64.
- [5] National Guideline Centre (UK). Evidence review for manual therapy for chronic primary pain. National Institute for Health and Care Excellence (NICE); 2021.
- [6] Keter DL, Bent JA, Bialosky JE, Courtney CA, Esteves JE, Funabashi M, et al. An international consensus on gaps in mechanisms of forced-based manipulation research: findings from a nominal group technique. *J Man Manip Ther* 2024;32:111–7.
- [7] Draper-Rodi J, Newell D, Barbe MF, Bialosky J. Integrated manual therapies: IASP taskforce viewpoint. *Pain Rep* 2024;9:e1192.
- [8] Roura S, Álvarez G, Solà I, Cerritelli F. Do manual therapies have a specific autonomic effect? An overview of systematic reviews. *PLoS One* 2021;16:e0260642.
- [9] Smith AR Jr. Manual therapy: the historical, current, and future role in the treatment of pain. *ScientificWorldJournal* 2007;7:109–20.
- [10] Oostendorp RAB. Credibility of manual therapy is at stake “Where do we go from here?” *J Man Manip Ther* 2018;26:189–92.

- [11] Lederman E. The fall of the postural-structural-biomechanical model in manual and physical therapies: Exemplified by lower back pain. *Journal of Bodywork and Movement Therapies* 2011;15:131–8. <https://doi.org/10.1016/j.jbmt.2011.01.011>.
- [12] Bialosky JE, Bishop MD, Penza CW. Placebo Mechanisms of Manual Therapy: A Sheep in Wolf’s Clothing? *J Orthop Sports Phys Ther* 2017;47:301–4.
- [13] Bialosky JE, Beneciuk JM, Bishop MD, Coronado RA, Penza CW, Simon CB, et al. Unraveling the Mechanisms of Manual Therapy: Modeling an Approach. *J Orthop Sports Phys Ther* 2018;48:8–18.
- [14] Gyer G, Michael J, Inklebarger J, Tedla JS. Spinal manipulation therapy: Is it all about the brain? A current review of the neurophysiological effects of manipulation. *J Integr Med* 2019;17:328–37.
- [15] D’Alessandro G, Cerritelli F, Cortelli P. Sensitization and Interoception as Key Neurological Concepts in Osteopathy and Other Manual Medicines. *Front Neurosci* 2016;10:100.
- [16] Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. *BMJ* 2021;374:n2061.
- [17] Hawe P, Shiell A, Riley T. Complex interventions: how “out of control” can a randomised controlled trial be? *BMJ* 2004;328:1561–3.
- [18] Wade DT, Smeets RJEM, Verbunt JA. Research in rehabilitation medicine: methodological challenges. *J Clin Epidemiol* 2010;63:699–704.
- [19] Kerry R. Expanding our perspectives on research in musculoskeletal science and practice. *Musculoskelet Sci Pract* 2017;32:114–9.
- [20] Dieppe P. Complex interventions. *Musculoskeletal Care* 2004;2:180–6.
- [21] Ford I, Norrie J. Pragmatic Trials. *New England Journal of Medicine* 2016;375:454–63. <https://doi.org/10.1056/nejmra1510059>.
- [22] Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, et al. Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ* 2008;337:a1655.
- [23] Bennell KL, Egerton T, Pua Y-H, Abbott JH, Sims K, Buchbinder R. Building the

rationale and structure for a complex physical therapy intervention within the context of a clinical trial: a multimodal individualized treatment for patients with hip osteoarthritis. *Phys Ther* 2011;91:1525–41.

[24] Craig P, Diepe P, Macintyre S. Medical Research Council (MRC) developing and evaluating complex interventions. London: MRC n.d.

[25] D'Alessandro G, Ruffini N, Iacopini A, Annoni M, Kossowsky J, Cerritelli F. Overcoming placebo-related challenges in manual therapy trials: The “whats and hows” and the “touch equality assumption” proposals. *Int J Osteopath Med* 2021;42:5–10.

[26] Esteves JE, Spence C. Diagnostic palpation and decision making: a neurocognitive model of expertise. *Russian Osteopathic J* 2014:1–2.

[27] Geri T, Viceconti A, Minacci M, Testa M, Rossettini G. Manual therapy: Exploiting the role of human touch. *Musculoskeletal Science and Practice* 2019;44:102044.

[28] Ezzatvar Y, Dueñas L, Balasch-Bernat M, Lluch-Girbés E, Rossettini G. Which portion of physiotherapy treatments' effect is attributable to contextual effects in people with musculoskeletal pain?: A meta-analysis of randomised placebo-controlled trials. *J Orthop Sports Phys Ther* 2024:1–28.

[29] Rossettini G, Camerone EM, Carlino E, Benedetti F, Testa M. Context matters: the psychoneurobiological determinants of placebo, nocebo and context-related effects in physiotherapy. *Arch Physiother* 2020;10:11.

[30] Rossettini G, Carlino E, Testa M. Clinical relevance of contextual factors as triggers of placebo and nocebo effects in musculoskeletal pain. *BMC Musculoskelet Disord* 2018;19:27.

[31] Rossettini G, Colombi A, Carlino E, Manoni M, Mirandola M, Polli A, et al. Unraveling negative expectations and nocebo-related effects in musculoskeletal pain. *Front Psychol* 2022;13. <https://doi.org/10.3389/fpsyg.2022.789377>.

[32] Hohenschurz-Schmidt D, Kleykamp BA, Draper-Rodi J, Vollert J, Chan J, Ferguson M, et al. Pragmatic trials of pain therapies: a systematic review of methods. *Pain* 2021.

[33] Fennelly O, Blake C, Desmeules F, Stokes D, Cunningham C. Patient-reported outcome measures in advanced musculoskeletal physiotherapy practice: a systematic

review. *Musculoskeletal Care* 2018;16:188–208.

[34] Hamilton DF, Giesinger JM, Giesinger K. It is merely subjective opinion that patient-reported outcome measures are not objective tools. *Bone Joint Res* 2017;6:665–6.

[35] Qian Y, Walters SJ, Jacques R, Flight L. Comprehensive review of statistical methods for analysing patient-reported outcomes (PROs) used as primary outcomes in randomised controlled trials (RCTs) published by the UK's Health Technology Assessment (HTA) journal (1997-2020). *BMJ Open* 2021;11:e051673.

[36] Pe M, Dorme L, Coens C, Basch E, Calvert M, Campbell A, et al. Statistical analysis of patient-reported outcome data in randomised controlled trials of locally advanced and metastatic breast cancer: a systematic review. *Lancet Oncol* 2018;19:e459–69.

[37] Alvarez G, Roura S, Cerritelli F, Esteves JE, Verbeeck J, van Dun PLS. The Spanish Osteopathic Practitioners Estimates and Rates (OPERA) study: A cross-sectional survey. *PLoS One* 2020;15:e0234713.

[38] Kamper SJ, Moseley AM, Herbert RD, Maher CG, Elkins MR, Sherrington C. 15 years of tracking physiotherapy evidence on PEDro, where are we now? *Br J Sports Med* 2015;49:907–9.

[39] Godwin M, Ruhland L, Casson I, MacDonald S, Delva D, Birtwhistle R, et al. Pragmatic controlled clinical trials in primary care: the struggle between external and internal validity. *BMC Med Res Methodol* 2003;3:28.

[40] Burns PB, Rohrich RJ, Chung KC. The levels of evidence and their role in evidence-based medicine. *Plast Reconstr Surg* 2011;128:305–10.

[41] Patsopoulos NA. A pragmatic view on pragmatic trials. *Dialogues Clin Neurosci* 2011;13:217–24.

[42] Kaptchuk TJ. The double-blind, randomized, placebo-controlled trial: gold standard or golden calf? *J Clin Epidemiol* 2001;54:541–9.

[43] Huebschmann AG, Leavitt IM, Glasgow RE. Making Health Research Matter: A Call to Increase Attention to External Validity. *Annu Rev Public Health* 2019;40:45–63.

[44] Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. *J Chronic Dis* 1967;20:637–48.

- [45] Singal AG, Higgins PDR, Waljee AK. A primer on effectiveness and efficacy trials. *Clin Transl Gastroenterol* 2014;5:e45.
- [46] Thorpe KE, Zwarenstein M, Oxman AD, Treweek S, Furberg CD, Altman DG, et al. A pragmatic–explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *Journal of Clinical Epidemiology* 2009;62:464–75. <https://doi.org/10.1016/j.jclinepi.2008.12.011>.
- [47] Treweek S, Zwarenstein M. Making trials matter: pragmatic and explanatory trials and the problem of applicability. *Trials* 2009;10:37.
- [48] Daniel Maddox C, Subialka JA, Young JL, Rhon DI. TITLE: Over Half of Clinical Trials of Mobilization and Manipulation for Patients with Low Back Pain May Have Limited Real-World Applicability. A Systematic Review of 132 Clinical Trials n.d.
- [49] Glasziou P, Matthews R, Boutron I, Chalmers I, Armitage P. The differences and overlaps between “explanatory” and “pragmatic” controlled trials: a historical perspective. *J R Soc Med* 2023;1410768231207536.
- [50] Nicholls SG, Zwarenstein M, Hey SP, Giraudeau B, Campbell MK, Taljaard M. The importance of decision intent within descriptions of pragmatic trials. *J Clin Epidemiol* 2020;125:30–7.
- [51] Porzsolt F, Wiedemann F, Phlippen M, Weiss C, Weiss M, Schmalin K, et al. The terminology conflict on efficacy and effectiveness in healthcare. *J Comp Eff Res* 2020;9:1171–8.
- [52] Fritz JM, Cleland J. Effectiveness versus efficacy: more than a debate over language. *J Orthop Sports Phys Ther* 2003;33:163–5.
- [53] Witt CM. Efficacy, effectiveness, pragmatic trials--guidance on terminology and the advantages of pragmatic trials. *Forsch Komplementmed* 2009;16:292–4.
- [54] Hohenschurz-Schmidt D, Vase L, Scott W, Annoni M, Ajayi OK, Barth J, et al. Recommendations for the development, implementation, and reporting of control interventions in efficacy and mechanistic trials of physical, psychological, and self-management therapies: the CoPPS Statement. *BMJ* 2023;381:e072108.
- [55] Walach H, Falkenberg T, Fønnebø V, Lewith G, Jonas WB. Circular instead of hierarchical: methodological principles for the evaluation of complex interventions. *BMC*

Med Res Methodol 2006;6:29.

[56] Minary L, Trompette J, Kivits J, Cambon L, Tarquinio C, Alla F. Which design to evaluate complex interventions? Toward a methodological framework through a systematic review. BMC Med Res Methodol 2019;19:92.

[57] Pinnock H, Epiphaniou E, Sheikh A, Griffiths C, Eldridge S, Craig P, et al. Developing standards for reporting implementation studies of complex interventions (StaRI): a systematic review and e-Delphi. Implement Sci 2015;10:42.

[58] Koes BW. How to evaluate manual therapy: value and pitfalls of randomized clinical trials. Man Ther 2004;9:183–4.

[59] Cashin AG, Lee H, Bagg MK, O'Hagan E, Traeger AC, Kamper SJ, et al. A systematic review highlights the need to improve the quality and applicability of trials of physical therapy interventions for low back pain. J Clin Epidemiol 2020;126:106–15.

[60] Riley SP, Swanson B, Brismée J-M, Sawyer SF. A systematic review of orthopaedic manual therapy randomized clinical trials quality. Journal of Manual & Manipulative Therapy 2016;24:241–52.
<https://doi.org/10.1080/10669817.2015.1119372>.

[61] Riley SP, Swanson BT, Sawyer SF, Brismée J-M. Is research quality in orthopedic manual therapy trials stagnating? Reflections and pathways for improving research quality and advance our profession. J Man Manip Ther 2016;24:239–40.

[62] Moseley AM, Elkins MR, Janer-Duncan L, Hush JM. The Quality of Reports of Randomized Controlled Trials Varies between Subdisciplines of Physiotherapy. Physiother Can 2014;66:36–43.

[63] Gonzalez GZ, Moseley AM, Maher CG, Nascimento DP, Costa L da CM, Costa LO. Methodologic Quality and Statistical Reporting of Physical Therapy Randomized Controlled Trials Relevant to Musculoskeletal Conditions. Arch Phys Med Rehabil 2018;99:129–36.

[64] Clar C, Tsertsivadze A, Court R, Hundt GL, Clarke A, Sutcliffe P. Clinical effectiveness of manual therapy for the management of musculoskeletal and non-musculoskeletal conditions: systematic review and update of UK evidence report. Chiropr Man Therap 2014;22:12.

- [65] Innocenti T, Giagio S, Salvioli S, Feller D, Minnucci S, Brindisino F, et al. Completeness of Reporting Is Suboptimal in Randomized Controlled Trials Published in Rehabilitation Journals, With Trials With Low Risk of Bias Displaying Better Reporting: A Meta-research Study. *Archives of Physical Medicine and Rehabilitation* 2022. <https://doi.org/10.1016/j.apmr.2022.01.156>.
- [66] Fregni F, Imamura M, Chien HF, Lew HL, Boggio P, Kaptchuk TJ, et al. Challenges and recommendations for placebo controls in randomized trials in physical and rehabilitation medicine: a report of the international placebo symposium working group. *Am J Phys Med Rehabil* 2010;89:160–72.
- [67] Levack WM, Malmivaara A, Meyer T, Negrini S. Methodological problems in rehabilitation research. Report from a cochrane rehabilitation methodology meeting. *Eur J Phys Rehabil Med* 2019;55:319–21.
- [68] Hart T. Treatment definition in complex rehabilitation interventions. *Neuropsychol Rehabil* 2009;19:824–40.
- [69] Degenhardt B, van Dun PLS, Jacobson E, Fritz S, Mettler P, Kettner N, et al. Profession-based manual therapy nomenclature: exploring history, limitations, and opportunities. *J Man Manip Ther* 2024;32:96–110.
- [70] Nim CG, Downie A, O'Neill S, Kawchuk GN, Perle SM, Leboeuf-Yde C. The importance of selecting the correct site to apply spinal manipulation when treating spinal pain: Myth or reality? A systematic review. *Sci Rep* 2021;11:23415.
- [71] Gorgos KS, Wasylyk NT, Van Lunen BL, Hoch MC. Inter-clinician and intra-clinician reliability of force application during joint mobilization: a systematic review. *Man Ther* 2014;19:90–6.
- [72] Wholohan AJ, Jedynek T. Intraclinician variability in force application during anteroposterior mobilization of the ankle joint. *J Chiropr Med* 2022;21:9–14.
- [73] Cook CE, Thigpen CA. Five good reasons to be disappointed with randomized trials. *J Man Manip Ther* 2019;27:63–5.
- [74] Cholewicki J, Pathak P, Reeves NP, Popovich JM Jr. Simulation of Multifactorial Causes of Low Back Pain. *Spine J* 2016;16:S277.
- [75] Lowe DB, Taylor MJ, Hill SJ. Cross-sectional examination of musculoskeletal

conditions and multimorbidity: influence of different thresholds and definitions on prevalence and association estimates. *BMC Res Notes* 2017;10:51.

[76] Menzel NN. Psychosocial factors in musculoskeletal disorders. *Crit Care Nurs Clin North Am* 2007;19:145–53.

[77] D'Alessandro G, Ruffini N, Iacopini A, Annoni M, Kossowsky J, Cerritelli F. Five challenges for manual therapies trials with placebo controls: A proposal. *Int J Osteopath Med* 2022;46:55–9.

[78] D'Alessandro G, Ruffini N, Aquino A, Galli M, Innocenti M, Tramontano M, et al. Differences between experimental and placebo arms in manual therapy trials: a methodological review. *BMC Med Res Methodol* 2022;22:219.

[79] Hettinga DM, Hurley DA, Jackson A, May S, Mercer C, Roberts L. Assessing the effect of sample size, methodological quality and statistical rigour on outcomes of randomised controlled trials on mobilisation, manipulation and massage for low back pain of at least 6 weeks duration. *Physiotherapy* 2008;94:97–104.

[80] Alvarez G, Núñez-Cortés R, Solà I, Sitjà-Rabert M, Fort-Vanmeerhaeghe A, Fernández C, et al. Sample size, study length, and inadequate controls were the most common self-acknowledged limitations in manual therapy trials: A methodological review. *J Clin Epidemiol* 2021;130:96–106.

[81] Dal-Ré R, Bracken MB, Ioannidis JPA. Call to improve transparency of trials of non-regulated interventions. *BMJ* 2015;350:h1323.

[82] Boutron I, Guittet L, Estellat C, Moher D, Hróbjartsson A, Ravaud P. Reporting methods of blinding in randomized trials assessing nonpharmacological treatments. *PLoS Med* 2007;4:e61.

[83] Malmivaara A. Pure intervention effect or effect in routine health care - blinded or non-blinded randomized controlled trial. *BMC Med Res Methodol* 2018;18:91.

[84] Sackett DL. Clinician-trialist rounds: 6. Testing for blindness at the end of your trial is a mug's game. *Clin Trials* 2011;8:674–6.

[85] Greenhalgh T, Howick J, Maskrey N, Evidence Based Medicine Renaissance Group. Evidence based medicine: a movement in crisis? *BMJ* 2014;348:g3725.

[86] Turrillas M, Sitjà-Rabert M, Pardo H, Vilaró Casamitjana J, Fort-Vanmeerhaeghe

A, Morral Fernández A, et al. Identification and description of controlled clinical trials published in Physiotherapy journals in Spain. *J Eval Clin Pract* 2017;23:29–36.

[87] Hanson K, Briki N, Erlangga D, Alebachew A, De Allegri M, Balabanova D, et al. The Lancet Global Health Commission on financing primary health care: putting people at the centre. *Lancet Glob Health* 2022;10:e715–72.

[88] Mayor E, Pérez G, Sánchez BS, Barco Á. Dificultades para el desarrollo de proyectos de investigación públicos en atención primaria. *Atencion Primaria* 2008;40:536–7.

[89] Deaton A, Cartwright N. Understanding and misunderstanding randomized controlled trials. *Soc Sci Med* 2018;210:2–21.

[90] Bird L, Arthur A, Cox K. “Did the trial kill the intervention?” experiences from the development, implementation and evaluation of a complex intervention. *BMC Med Res Methodol* 2011;11:24.

[91] Hohenschurz-Schmidt D, Draper-Rodi J, Vase L, Scott W, McGregor A, Soliman N, et al. Blinding and Sham Control Methods in Trials of Physical, Psychological, and Self-Management Interventions for Pain (Article I): a Systematic Review and Description of Methods. *Pain* 2022;Publish Ahead of Print. <https://doi.org/10.1097/j.pain.0000000000002723>.

[92] Álvarez, G Tesis doctoral n.d. <http://hdl.handle.net/10803/674940> (accessed October 24, 2024).

[93] Calvert M, Blazeby J, Altman DG, Revicki DA, Moher D, Brundage MD, et al. Reporting of patient-reported outcomes in randomized trials: The CONSORT PRO extension. *JAMA* 2013;309:814.

[94] Rothwell PM. External validity of randomised controlled trials: “To whom do the results of this trial apply?” *The Lancet* 2005;365:82–93. [https://doi.org/10.1016/s0140-6736\(04\)17670-8](https://doi.org/10.1016/s0140-6736(04)17670-8).

[95] Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. *BMJ* 2015;350:h2147.

[96] Correction Chalmers I, Bracken MB, Djulbegovic B, Garattini S, Grant J, Gülmezoglu AM, et al. How to increase value and reduce waste when research priorities

are set. *Lancet* 2014;383:156–65.

[97] Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, et al. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *BMJ* 2008;337:a2390.

[98] Dal-Ré R, Janiaud P, Ioannidis JPA. Real-world evidence: How pragmatic are randomized controlled trials labeled as pragmatic? *BMC Med* 2018;16:49.

[99] Willis A, Shiely F, Treweek S, Taljaard M, Loudon K, Howie A, et al. Comments, suggestions and criticisms of the PRECIS-2 design tool: A citation analysis. *J Clin Epidemiol* 2024;111:534.

[100] Johnson KE, Neta G, Dember LM, Coronado GD, Suls J, Chambers DA, et al. Use of PRECIS ratings in the National Institutes of Health (NIH) Health Care Systems Research Collaboratory. *Trials* 2016;17:32.

[101] Darker C, Loudon K, O’Connell N, Castello S, Burke E, Vance J, et al. An application of PRECIS-2 to evaluate trial design in a pilot cluster randomised controlled trial of a community-based smoking cessation intervention for women living in disadvantaged areas of Ireland. *Pilot Feasibility Stud* 2022;8:19.

[102] Sepehrvand N, Alemayehu W, Das D, Gupta AK, Gouda P, Ghimire A, et al. Trends in the Explanatory or Pragmatic Nature of Cardiovascular Clinical Trials Over 2 Decades. *JAMA Cardiol* 2019;4:1122–8.

[103] Devos F, Foissac F, Bouazza N, Ancel P-Y, Tréluyer J-M, Chappuy H. Study characteristics impacted the pragmatism of randomized controlled trial published in nursing: a meta-epidemiological study. *J Clin Epidemiol* 2019;116:18–25.

[104] Rosas LG, Lv N, Azar K, Xiao L, Yank V, Ma J. Applying the Pragmatic–Explanatory Continuum Indicator Summary Model in a Primary Care–Based Lifestyle Intervention Trial. *Am J Prev Med* 2015;49:S208–14.

[105] Nguyen HQ, Moy ML, Fan VS, Gould MK, Xiang A, Bailey A, et al. Applying the pragmatic-explanatory continuum indicator summary to the implementation of a physical activity coaching trial in chronic obstructive pulmonary disease. *Nurs Outlook* 2018;66:455–63.

[106] Elder WG, Munk N. Using the Pragmatic-Explanatory Continuum Indicator

Summary (PRECIS) Model in Clinical Research: Application to Refine a Practice-based Research Network (PBRN) Study. *The Journal of the American Board of Family Medicine* 2014;27:846–54. <https://doi.org/10.3122/jabfm.2014.06.140042>.

[107] Riddle DL, Johnson RE, Jensen MP, Keefe FJ, Kroenke K, Bair MJ, et al. The Pragmatic-Explanatory Continuum Indicator Summary (PRECIS) instrument was useful for refining a randomized trial design: Experiences from an investigative team. *Journal of Clinical Epidemiology* 2010;63:1271–5. <https://doi.org/10.1016/j.jclinepi.2010.03.006>.

[108] Palakshappa JA, Gibbs KW, Lannan MT, Cranford AR, Taylor SP. Systematic Review of the “Pragmatism” of Pragmatic Critical Care Trials. *Critical Care Explorations* 2022;4:e0738.

[109] Taljaard M, Nicholls SG, Howie AH, Nix HP, Carroll K, Moon PM, et al. An analysis of published trials found that current use of pragmatic trial labels is uninformative. *Journal of Clinical Epidemiology* 2022. <https://doi.org/10.1016/j.jclinepi.2022.08.007>.

[110] Neta G, Johnson KE. Informing real-world practice with real-world evidence: the value of PRECIS-2. *BMC Med* 2018;16:76.

[111] Dal-Ré R, de Boer A, James SK. The design can limit PRECIS-2 retrospective assessment of the clinical trial explanatory/pragmatic features. *J Clin Epidemiol* 2020;126:193–201.

[112] Leviton LC. Generalizing about Public Health Interventions: A Mixed-Methods Approach to External Validity. *Annu Rev Public Health* 2017;38:371–91.

[113] Ernst E, Canter PH. Limitations of “pragmatic” trials. *Postgrad Med J* 2005;81:203–203.

[114] Kent DM, Kitsios G. Against pragmatism: on efficacy, effectiveness and the real world. *Trials* 2009;10:48.

[115] Rowbotham MC, Gilron I, Glazer C, Rice ASC, Smith BH, Stewart WF, et al. Can pragmatic trials help us better understand chronic pain and improve treatment? *Pain* 2013;154:643–6.

[116] Zwarenstein M, Thorpe K, Treweek S, Loudon K. PRECIS-2 for retrospective assessment of RCTs in systematic reviews. *J Clin Epidemiol* 2020;126:202–6.

[117] Dal-Ré R. The PRECIS-2 tool seems not to be useful to discriminate the degree

of pragmatism of medicine masked trials from that of open-label trials. *Eur J Clin Pharmacol* 2021;77:539–46.

[118] Dal-Ré R. Pragmatic trials, blinding, placebos, and the usefulness of the PRECIS-2 tool. *Eur J Clin Pharmacol* 2021;77:1071–2.

[119] Zwarenstein M, Howie A. Blinding, pragmatism, and the PRECIS-2 tool for designing and assessing randomized trials. *Eur J Clin Pharmacol* 2021;77:1069–70.

[120] Gamerman V, Cai T, Elsässer A. Pragmatic randomized clinical trials: best practices and statistical guidance. *Health Serv Outcomes Res Methodol* 2019;19:23–35.

[121] Wieland LS, Berman BM, Altman DG, Barth J, Bouter LM, D’Adamo CR, et al. Rating of Included Trials on the Efficacy-Effectiveness Spectrum: development of a new tool for systematic reviews. *J Clin Epidemiol* 2017;84:95–104.

[122] Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. *Cochrane handbook for systematic reviews of interventions*. 2nd ed. Standards Information Network; 2019.

[123] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Rev Esp Cardiol* 2021;74:790–9.

[124] Sacristán JA, Dilla T. Pragmatic trials revisited: applicability is about individualization. *J Clin Epidemiol* 2018;99:164–6.

[125] Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. *Syst Rev* 2016;5:210.

[126] Zhang Y-Q, Jiao R-M, Witt CM, Lao L, Liu J-P, Thabane L, et al. How to design high quality acupuncture trials—a consensus informed by evidence. *BMJ* 2022;376. <https://doi.org/10.1136/bmj-2021-067476>.

[127] Boutron I, Altman DG, Moher D, Schulz KF, Ravaud P, CONSORT NPT Group. CONSORT Statement for Randomized Trials of Nonpharmacologic Treatments: A 2017 Update and a CONSORT Extension for Nonpharmacologic Trial Abstracts. *Ann Intern Med* 2017;167:40–7.

[128] Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel

group randomised trials. *BMJ* 2010;340:c869.

[129] Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;343:d5928.

[130] Cross J, Elender F, Barton G, Clark A, Shepstone L, Blyth A, et al. A randomised controlled equivalence trial to determine the effectiveness and cost-utility of manual chest physiotherapy techniques in the management of exacerbations of chronic obstructive pulmonary disease (MATREX). *Health Technology Assessment* 2010;14. <https://doi.org/10.3310/hta14230>.

[131] Alvarez G, Cerritelli F, Urrutia G. Using the template for intervention description and replication (TIDieR) as a tool for improving the design and reporting of manual therapy interventions. *Man Ther* 2016;24:85–9.

[132] Janiaud P, Dal-Ré R, Ioannidis JPA. Assessment of Pragmatism in Recently Published Randomized Clinical Trials. *JAMA Internal Medicine* 2018;178:1278. <https://doi.org/10.1001/jamainternmed.2018.3321>.

[133] Averitt AJ, Ryan PB, Weng C, Perotte A. A conceptual framework for external validity. *J Biomed Inform* 2021;121:103870.

[134] Nguyen AP, Pitance L, Mahaudens P, Detrembleur C, David Y, Hall T, et al. Effects of Mulligan Mobilization with Movement in Subacute Lateral Ankle Sprains: A Pragmatic Randomized Trial. *J Man Manip Ther* 2021;29:341–52.

[135] Mafetoni RR, Shimo AKK. Effects of acupressure on progress of labor and cesarean section rate: randomized clinical trial. *Rev Saude Publica* 2015;49:9.

[136] Wartolowska KA, Hohenschurz-Schmidt D, Vase L, Aronson JK. The importance of using placebo controls in nonpharmacological randomised trials. *Pain* 2023;164:921–5.

[137] Lin I, Wiles L, Waller R, Goucke R, Nagree Y, Gibberd M, et al. What does best practice care for musculoskeletal pain look like? Eleven consistent recommendations from high-quality clinical practice guidelines: systematic review. *Br J Sports Med* 2020;54:79–86.

[138] Gordon KS, Peduzzi P, Kerns RD. Designing Trials with Purpose: Pragmatic

Clinical Trials of Nonpharmacological Approaches for Pain Management. *Pain Med* 2020;21:S7–12.

[139] Negrini S, Arienti C, Pollet J, Engkasan JP, Francisco GE, Frontera WR, et al. Clinical replicability of rehabilitation interventions in randomized controlled trials reported in main journals is inadequate. *J Clin Epidemiol* 2019;114:108–17.

[140] Cottone KA, Schumacher MR, Young JL, Rhon DI. The majority of clinical trials assessing mobilization and manipulation for neck pain lack a pragmatic approach: a systematic review of 174 trials. *J Man Manip Ther* 2024;32:478–94.

[141] Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ* 2014;348:g1687.

[142] Hohenschurz-Schmidt DJ, Cherkin D, Rice ASC, Dworkin RH, Turk DC, McDermott MP, et al. Research objectives and general considerations for pragmatic clinical trials of pain treatments: IMMPACT statement. *Pain* 2023. <https://doi.org/10.1097/j.pain.0000000000002888>.

[143] Sacristán JA, Knottnerus JA. Individual point-of-care trials: a new approach to incorporate patient's preferences into personalized pragmatic clinical trials. *J Clin Epidemiol* 2021;130:152–5.

[144] Roura, Alvarez, Hohenschurz-Schmidt D, Solà, Núñez-Cortés R, Bracchiglione, et al. A call for improving the Internal validity and the reporting of manual therapy trials self-labelled as pragmatic: A methodological review. *Int J Osteopath Med* 2025:100754.

[145] Roura S, Alvarez G, Hohenschurz-Schmidt D, Solà I, Núñez-Cortés R, Bracchiglione J, et al. Lack of pragmatic attitude of self-labelled pragmatic trials on manual therapy: a methodological review. *BMC Med Res Methodol* 2024;24. <https://doi.org/10.1186/s12874-024-02393-1>.

[146] Leech JB, Owen WE, Young JL, Rhon DI. Incomplete reporting of manual therapy interventions and a lack of clinician and setting diversity in clinical trials for neck pain limits replication and real-world translation. A scoping review. *Journal of Manual & Manipulative Therapy* 2022;1–10. <https://doi.org/10.1080/10669817.2022.2113295>.

[147] Hoffmann TC, Erueti C, Glasziou PP. Poor description of non-pharmacological

interventions: analysis of consecutive sample of randomised trials. *BMJ* 2013;347:f3755–f3755. <https://doi.org/10.1136/bmj.f3755>.

[148] Glasziou P, Meats E, Heneghan C, Shepperd S. What is missing from descriptions of treatment in trials and reviews? *BMJ* 2008;336:1472–4.

[149] Dijkers MP. Overview of Reviews Using the Template for Intervention Description and Replication (TIDieR) as a Measure of Trial Intervention Reporting Quality. *Arch Phys Med Rehabil* 2021;102:1623–32.

[150] Yamato T, Maher C, Saragiotto B, Moseley A, Hoffmann T, Elkins M, et al. Improving completeness and transparency of reporting in clinical trials using the template for intervention description and replication (TIDieR) checklist will benefit the physiotherapy profession. *Journal of Manual & Manipulative Therapy* 2016;24:183–4. <https://doi.org/10.1080/10669817.2016.1210343>.

[151] Gastaldon C, Mosler F, Toner S, Tedeschi F, Bird VJ, Barbui C, et al. Are trials of psychological and psychosocial interventions for schizophrenia and psychosis included in the NICE guidelines pragmatic? A systematic review. *PLoS One* 2019;14:e0222891.

[152] Arienti C, Armijo-Olivo S, Ferriero G, Feys P, Hoogeboom T, Kiekens C, et al. The influence of bias in randomized controlled trials on rehabilitation intervention effect estimates: what we have learned from meta-epidemiological studies. *Eur J Phys Rehabil Med* 2024;60:135–44.

[153] Page MJ, Higgins J, Clayton G, Sterne J. Hróbjartsson , Savović J. Empirical evidence study design biases randomized trials: systematic review meta-epidemiological studies. *PLoS One* 2016;11.

[154] Moustgaard H, Clayton GL, Jones HE, Boutron I, Jørgensen L, Laursen DRT, et al. Impact of blinding on estimated treatment effects in randomised clinical trials: meta-epidemiological study. *BMJ* 2020;368:l6802.

[155] Armijo-Olivo S, Dennett L, Arienti C, Dahchi M, Arokoski J, Heinemann AW, et al. Blinding in rehabilitation research: Empirical evidence on the association between blinding and treatment effect estimates. *Am J Phys Med Rehabil* 2020;99:198–209.

[156] Villamar MF, Contreras VS, Kuntz RE, Fregni F. The reporting of blinding in physical medicine and rehabilitation randomized controlled trials: a systematic review. *J Rehabil*

Med 2013;45:6–13.

[157] Kyte DG, Calvert M, van der Wees PJ, ten Hove R, Tolan S, Hill JC. An introduction to patient-reported outcome measures (PROMs) in physiotherapy. *Physiotherapy* 2015;101:119–25.

[158] Dal-Ré R, Rosendaal F. Efficacy and effectiveness: The wrong use of different terms. *Eur J Intern Med* 2018;54:e17–8.

[159] Schmalting K, Kaplan RM, Porzsolt F. Efficacy and effectiveness studies of depression are not well-differentiated in the literature: a systematic review. *BMJ Evid Based Med* 2021;26:28–30.

[160] Park S-Y, Hwang E-H, Cho J-H, Kim K-W, Ha I-H, Kim M-R, et al. Comparative Effectiveness of Chuna Manipulative Therapy for Non-Acute Lower Back Pain: A Multi-Center, Pragmatic, Randomized Controlled Trial. *J Clin Med Res* 2020;9. <https://doi.org/10.3390/jcm9010144>.

[161] Adams KR, Famuyide AO, Young JL, Maddox CD, Rhon DI. Pragmatism in manual therapy trials for knee osteoarthritis: a systematic review. *Arch Physiother* 2024;14:1–10.

[162] Núñez-Cortés R, Alvarez G, Pérez-Bracchiglione J, Cabanas-Valdés R, Calvo-Sanz J, Bonfill X, et al. Reporting results in manual therapy clinical trials: A need for improvement. *Int J Osteopath Med* 2021;0. <https://doi.org/10.1016/j.ijosm.2021.06.002>.

[163] Roura S, Alvarez G, Hohenschurz-Schmidt D, Solà I, Núñez-Cortés R, Bracchiglione J, et al. Alignment between the research question, design and terminology is required in manual therapy trials - a methodological study. *J Clin Epidemiol* 2025;183:111817.

[164] Hohenschurz-Schmidt D, Draper-Rodi DJ, Vase PL, Scott DW, McGregor PA, Soliman DN, et al. Blinding and sham control methods in trials of physical, psychological, and self-management interventions for pain (article II): a meta-analysis relating methods to trial results. *Pain* 2022;10.1097/j.pain.0000000000002730.

[165] Sénéquier A, Draper-Rodi J, Alvarez Bustins G, Braithwaite F, Brown J, Corcoran D, et al. Investigating the trustworthiness of randomised controlled trials in osteopathic research: a systematic review with meta-analysis. *J Clin Epidemiol* 2025;183:111788.

