

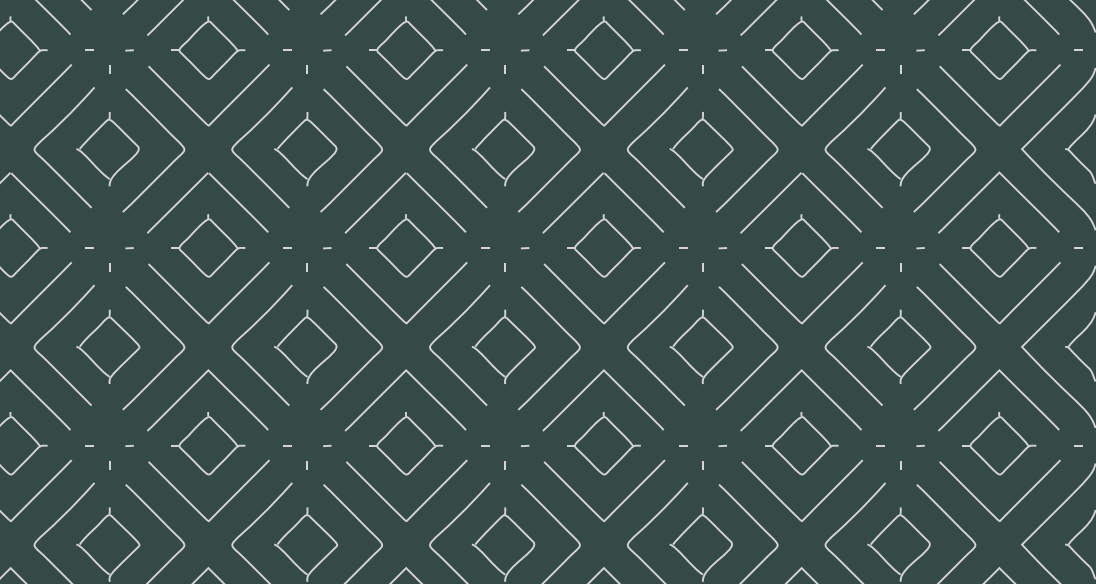
**Retos metodológicos en la
evaluación de la eficacia de
la terapia manual**

Gerard Alvarez Bustins

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:  http://cat.creativecommons.org/?page_id=184

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:  <http://es.creativecommons.org/blog/licencias/>

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

Retos metodológicos en la evaluación de la eficacia de la terapia manual

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

Gerard Alvarez Bustins

Directores: Dr. Gerard Urrútia Cuchí y Dr. Xavier Bonfill Cosp

Octubre 2021

UAB

Universitat Autònoma de Barcelona

TESIS DOCTORAL

RETOS METODOLÓGICOS EN LA EVALUACIÓN DE LA EFICACIA DE LA TERAPIA MANUAL

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

Gerard Alvarez Bustins

Directores: Dr. Gerard Urrútia Cuchí y
Dr. Xavier Bonfill Cosp

Octubre 2021



Universitat Autònoma de Barcelona

Agradecimientos

A Débora y Bruno. Todo esto es para ellos y gracias a ellos.

A mis directores de tesis, los Drs. Gerard Urrútia y Xavier Bonfill

A mis co-autores en las publicaciones de este trabajo de tesis.
Muy en especial a Rodrigo Nuñez e Ivan Solà.

A todas las personas que me han apoyado en este camino.
En especial a Sonia Roura

***“El éxito es el resultado de una
infatigable atención al propósito”***

Disraeli, estadista inglés del siglo IX

Resumen

Antecedentes

La Terapia Manual (TM) es un tratamiento físico basado en la aplicación de fuerzas y movimientos a través de las manos con una finalidad terapéutica. Los procesos de rehabilitación, dentro de los cuáles se engloban las terapias físicas y en particular la TM, responden a la denominación de intervenciones complejas (IC) por estar compuestas comúnmente por “paquetes terapéuticos” que pueden incluir varios abordajes y técnicas diferentes. Debido a sus particularidades, la evaluación científica de las IC a partir de diseños experimentales clásicos como el ensayo clínico aleatorizado (ECA), supone un reto metodológico. Entre otros factores, la dificultad de alcanzar un equilibrio razonable entre la validez interna y la validez externa de estos estudios, conlleva la realización de estudios considerados frecuentemente de baja calidad. A pesar del aumento en la producción científica en los últimos años, tanto la calidad metodológica como la calidad del reporte de los estudios en TM sigue siendo cuestionada. Todo ello evidencia la necesidad de encontrar mejores diseños de investigación adaptados a sus particularidades y a mejorar la calidad del reporte de las publicaciones. Para conseguir tal propósito, resulta imprescindible conocer de forma precisa cuáles son las dificultades específicas en este campo y establecer potenciales soluciones metodológicas que puedan conducir a mejorar la producción científica en el futuro.

Objetivos

Este trabajo de tesis tiene como objetivo analizar de forma exhaustiva cuáles son los elementos críticos que influyen de forma negativa en la calidad de la evidencia reportada en el campo de las TMs. A su vez, se analizan las dificultades metodológicas inherentes a la evaluación de la eficacia de este tipo de intervenciones y, finalmente, se apuntan estrategias de mejora, adaptaciones y alternativas en los diseños de investigación para conseguir una mayor calidad metodológica y del reporte.

Métodos

Se han llevado a cabo 3 trabajos de revisión a partir de una amplia muestra de ECAs en TM. En el primer trabajo, se evaluó la calidad del reporte de una muestra representativa de ECAs en TM. Esta evaluación del reporte se centró en aspectos específicos de la validez, la fiabilidad y el grado de detalle en la descripción de las intervenciones. A su vez, se compararon los resultados entre dos periodos de tiempo equivalentes - previo y posterior a la publicación de la extensión CONSORT para intervenciones no-farmacológicas (CONSORTnpt) - para evaluar el impacto de dicha guía en la calidad del reporte de los ECAs seleccionados. En el segundo trabajo, la muestra de estudios fue ampliada y se analizaron las limitaciones autoreportadas (LARs) por parte de los autores. Se realizó un análisis descriptivo de las LARs, se categorizaron y se discutieron

los motivos expuestos por parte de los autores para justificar su inclusión. En el tercer trabajo, se realizó un análisis exhaustivo centrado en el reporte de los resultados junto a un sub-análisis por franjas temporales de 5 años para valorar posibles mejoras en el tiempo. Esta evaluación consideró tanto el reporte de los resultados incluido en el texto como el reporte gráfico. Asimismo, se presenta un artículo adicional (como anexo) acerca de la pertinencia del uso de la guía Template for Intervention Description and Replication (TIDieR) en el campo de las TMs

Resultados

Los resultados del primer trabajo indican que la calidad general del reporte de ECAs en TM es deficiente y que este no ha mejorado con el tiempo, salvo en unos pocos aspectos. Además, los resultados sugieren que las guías de reporte (especialmente CONSORTnpt) no han tenido un impacto significativo en la calidad de los informes a pesar de los años transcurridos desde su publicación. En relación al segundo trabajo, nuestros resultados indican que al menos dos limitaciones diferentes se reconocen sistemáticamente en los ECAs de TM. En relación al tipo de limitaciones, las más comúnmente declaradas fueron las relacionadas con el tamaño de la muestra (47,5%), seguidas de las limitaciones relacionadas con la duración del estudio y el seguimiento (33,3%) y el uso de controles inadecuados (32,5%). Los resultados del tercer trabajo profundizaron en la información obtenida en el primero y constataron que el reporte de los resultados en ECAs de TM es generalmente incompleto, lo que limita de forma importante la correcta interpretación de los mismos. Se evidencia también una deficiente presentación gráfica de estos resultados siendo en muchos casos incompleta y/o carente de claridad visual.

Conclusiones

Nuestros hallazgos sugieren que los ECAs que evalúan intervenciones manuales son reportados de forma deficiente incluso años después de la disponibilidad de guías de reporte estandarizadas, lo que limita su potencial aplicabilidad en la práctica clínica. Las principales deficiencias en el reporte y limitaciones identificadas, se relacionan con los siguientes elementos metodológicos: i) el tamaño de las muestras, ii) las intervenciones control, iii) las intervenciones experimentales, iv) las estrategias de enmascaramiento, v) la duración del estudio y su seguimiento y vi) las medidas de resultado. Nuestros resultados también muestran que la presentación gráfica de los resultados a menudo es incompleta y carece de claridad visual. La existencia de déficits metodológicos en los ECAs en TM y la constatación de que estos siguen muy presentes a pesar del paso del tiempo, pone de relieve las dificultades que entraña el diseño y realización de estudios experimentales de alta calidad para evaluar la eficacia de una intervención compleja no-farmacológica como la TM

Resum

Antecedents

La Teràpia Manual (TM) és un tractament físic basat en l'aplicació de forces i moviments a través de les mans amb una finalitat terapèutica. Els processos de rehabilitació, dins dels quals s'engloben les teràpies físiques i en particular la TM, responen a la denominació d'intervencions complexes (IC) per estar compostes comunament per "paquets terapèutics" que poden incloure diversos abordatges i tècniques diferents. A causa de les seves particularitats, l'avaluació científica de les IC a partir de dissenys experimentals clàssics com l'assaig clínic aleatoritzat (ACA), suposa un repte metodològic. Entre d'altres factors, la dificultat d'arribar a un equilibri raonable entre la validesa interna i la validesa externa d'aquests estudis, comporta la realització d'estudis considerats freqüentment de baixa qualitat. Tot i l'augment en la producció científica en els darrers anys, tant la qualitat metodològica com la qualitat del raport dels estudis en TM segueix sent qüestionada. Tot plegat evidencia la necessitat de trobar millors dissenys d'investigació adaptats a les seves particularitats i a millorar la qualitat del raport de les publicacions. A tal efecte, és fa imprescindible conèixer de forma precisa quines són les dificultats específiques en aquest camp i establir potencials solucions metodològiques que puguin conduir a millorar la producció científica en el futur.

Objectius

Aquest treball de tesi té com a objectiu analitzar de forma exhaustiva quins són els elements crítics que influeixen de manera negativa en la qualitat de l'evidència reportada en el camp de les TMs. Al seu torn, s'analitzen les dificultats metodològiques inherents a l'avaluació de l'eficàcia d'aquest tipus d'intervencions i, finalment, s'apunten estratègies de millora, adaptacions i alternatives en els dissenys d'investigació per aconseguir una major qualitat metodològica i de l'informe.

Mètodes

S'han dut a terme 3 treballs de revisió a partir d'una àmplia mostra d'ACAs en TM. En el primer treball, es va avaluar la qualitat del raport d'una mostra representativa de ACAs en TM. Aquesta avaluació del raport es va centrar en aspectes específics de la validesa, la fiabilitat i el grau de detall en la descripció de les intervencions. Al seu torn, es van comparar els resultats entre dos períodes de temps equivalents - previ i posterior a la publicació de l'extensió CONSORT per a intervencions no-farmacològiques (CONSORTnpt) - per avaluar l'impacte d'aquesta guia en la qualitat del raport dels ACAs seleccionats. En el segon treball, la mostra d'estudis va ser ampliada i es van analitzar les limitacions auto-reportadas (LARs) per part dels autors. Es va realitzar una anàlisi descriptiva de les LARs, es van categoritzar i es van discutir els motius exposats per part dels autors per justificar la seva inclusió. En el tercer treball, es va realitzar una anàlisi

exhaustiva centrada en el report dels resultats així com un sub-anàlisi per franques temporals de 5 anys per valorar possibles millores en el temps. Aquesta avaluació va considerar tant el report dels resultats inclòs en el text com el report gràfic. Així mateix, es presenta un article addicional (com a annex) sobre la pertinència de l'ús de la guia Template for Intervention Description and Replication (TIDieR) en el camp de les TMs

Resultats

Els resultats del primer treball indiquen que la qualitat general del raport d'ACAs en TM és deficient i que aquest no ha millorat amb el temps, excepte en pocs aspectes. A més, els resultats suggereixen que les guies de raport (especialment CONSORTnpt) no han tingut un impacte significatiu en la qualitat del raport malgrat els anys transcorreguts des de la seva publicació. En relació a el segon treball, els nostres resultats indiquen que al menys dues limitacions diferents es reconeixen sistemàticament en els ACAs de TM. En relació al tipus de limitacions, les més comunament declarades van ser les relacionades amb la mida de la mostra (47,5%), seguides de les limitacions relacionades amb la durada de l'estudi i el seguiment (33,3%) i l'ús de controls inadequats (32,5%). Els resultats del tercer treball van aprofundir en la informació obtinguda en el primer i van constatar que el report dels resultats en ACAs de TM és generalment incomplet, el que limita de forma important la correcta interpretació dels mateixos. S'evidencia també una deficient presentació gràfica d'aquests resultats que en molts casos es incompleta i/o sense claredat visual.

Conclusions

Les nostres troballes suggereixen que els ACAs que avaluen intervencions manuals són reportats de forma deficient fins i tot anys després de la disponibilitat de guies de raport estandarditzades, el que limita la seva potencial aplicabilitat en la pràctica clínica. Les principals deficiències en el report i limitacions identificades, es relacionen amb els següents elements metodològics: i) la mida de les mostres, ii) les intervencions control, iii) les intervencions experimentals, iv) les estratègies d'emascarament, v) la durada de l'estudi i el seu seguiment i vi) les mesures de resultat. Els nostres resultats també mostren que la presentació gràfica dels resultats sovint és incompleta i no té claredat visual. L'existència de dèficits metodològics en els ECAs en TM i la constatació que aquests segueixen molt presents malgrat el pas el temps, posa en relleu les dificultats que comporta el disseny i realització d'estudis experimentals d'alta qualitat per avaluar l'eficàcia d'una intervenció complexa no-farmacològica com la TM

Abstract

Background

Manual Therapy (MT) is a physical treatment based on the application of forces and movements through the hands with a therapeutic purpose. Rehabilitation processes, such as physical and manual therapies are considered complex interventions (CI) because they are commonly composed of “therapeutic packages” that can include several different approaches and techniques. Due to its particularities, the scientific evaluation of CI from classic experimental designs such as the randomized clinical trial (RCT), represents a methodological challenge. Among other factors, the difficulty of reaching a reasonable balance between internal validity and external validity means that these studies are often considered of low quality. Despite the increase in scientific production in recent years, both the methodological and reporting quality of MT studies continues to be questioned. Overall, there is a need to find better research designs adapted to MT particularities and also to improve the quality of the report of the publications. In order to achieve this purpose, it is essential to know precisely what the specific difficulties are in this field and to establish potential methodological solutions that may lead to improving scientific production in the future.

Objectives

This thesis work aims to exhaustively analyze which are the critical elements that negatively influence the quality of the evidence reported in the field of MTs. At the same time, the specific methodological difficulties in evaluating the efficacy of this type of intervention are analyzed and, finally, improvement strategies, adaptations and alternatives in the research designs are pointed out to achieve a higher methodological and reporting quality in the future.

Methods

Three review studies have been carried out based on a large sample of RCTs in MT. In the first work, the quality of the report of a representative sample of RCTs in MT was evaluated. This evaluation of the report focused on specific aspects of validity, reliability and the degree of detail in the description of the interventions. In turn, the results were compared between two equivalent periods of time - before and after the publication of the CONSORT extension for non-pharmacological interventions (CONSORTnpt) - to evaluate the impact of this guide on the quality of the report of the selected RCTs. In the second work, the study sample was expanded and the authors' self-reported limitations (LARs) were analyzed. A descriptive analysis of the LARs was carried out, they were categorized and the reasons presented by the authors to justify their inclusion were discussed. In the third work, an exhaustive analysis focused on the reporting of the results was carried out together with a sub-analysis by time periods of 5 years to assess possible improvements over time. This evaluation considered both the report of the results included in the text and the graphic report. Li-

kewise, an additional article is presented (as an annex) about the relevance of using the Template for Intervention Description and Replication (TIDieR) guide in the field of MTs.

Results

The results of the first study indicate that the quality of the report of MT RCTs is poor and that it has not improved over time, except in a few aspects. Furthermore, the results suggest that the reporting guidelines (especially CONSORT-npt) have not had a significant impact on the quality of the reports despite the years that have elapsed since their publication. In relation to the second work, our results indicate that at least two different limitations are systematically recognized in MT RCTs. Regarding the type of limitations, the most commonly acknowledged were those related to the sample size (47.5%), followed by limitations related to the duration of the study and follow-up (33.3%) and the use of inadequate controls (32.5%). The results of the third work deepened into the information obtained in the first and found that the reporting of the results in MT RCTs is generally incomplete, which significantly limits their correct interpretation. There is also evidence of a poor graphic presentation of these results, being in many cases incomplete and/or lacking in visual clarity.

Conclusions

Our findings suggest that RCTs evaluating manual interventions are poorly reported even years after the availability of standardized reporting guidelines, limiting their potential applicability in clinical practice. The main deficiencies in the report and the limitations identified are related to the following methodological elements: i) the size of the samples, ii) the control interventions, iii) the experimental interventions, iv) the masking strategies, v) the duration of the study and its follow-up and vi) outcome measures. Our results also show that the graphical presentation of the results is often incomplete and lacks visual clarity. The existence of methodological deficits in the RCTs in MT and the evidence that these are still very present over time, highlights the difficulties involved in designing and conducting high-quality experimental studies to evaluate the efficacy of a non-pharmacological complex intervention such as MT.

Índice

1. Introducción	16
1.1 Fundamentos de la Terapia Manual	16
1.1.1 Definición	16
1.1.2 Historia de la Terapia Manual en Occidente	16
1.1.3 Mecanismos de acción de la Terapia Manual	18
1.1.4 Literatura científica en Terapia Manual	19
1.2 Investigación en Terapia Manual	20
1.2.1 El Ensayo Clínico Aleatorizado en Terapia Manual	20
1.2.2 La Terapia Manual como una intervención compleja	26
1.2.3 Características de la investigación en Intervenciones complejas	27
1.2.4 Calidad de la evidencia en Fisioterapia y Terapia Manual	30
1.2.5 CONSORT y el reporte en intervenciones no-farmacológicas	31
1.3 Justificación de la Tesis	33
2. Objetivos	36
2.1 Objetivo general	36
2.2 Objetivos específicos	36
3. Métodos	40
3.1 Métodos del primer trabajo	40
3.1.1 Criterios de elegibilidad	40
3.1.2 Estrategia de búsqueda	40
3.1.3 Cálculo del tamaño de la muestra	40
3.1.4 Extracción de datos	41
3.1.5 Análisis de los datos	41
3.2 Métodos del segundo trabajo	41
3.2.1 Criterios de elegibilidad	42
3.2.2 Extracción de datos	42
3.2.3 Análisis de datos	42
3.3 Métodos del tercer trabajo	42
3.3.1 Criterios de elegibilidad	42
3.3.2 Evaluación del reporte	43
3.3.4 Análisis de datos	44
4. Resultados	48
4.1 Resultados del primer trabajo	49
4.1.1 Resultados de la búsqueda	49
4.1.2 Descripción de los artículos	49
4.1.3 Análisis comparativo del reporte	50
4.1.4 Primera publicación	50

4.2 Resultados del segundo trabajo	66
4.2.1 Descripción de los artículos	66
4.2.2 Número de limitaciones	66
4.2.3 Categorías y limitaciones específicas reportadas	66
4.2.4 Segunda publicación	66
4.3 Resultados del tercer trabajo	80
4.3.1 Descripción de los artículos	80
4.3.2 Análisis del reporte de resultados en el texto	80
4.3.3 Análisis del reporte gráfico de resultados	80
4.3.4 Tercera publicación	81
5. Discusión	92
5.1 Principales resultados derivados de los trabajos	92
5.2 Discusión de los elementos más relevantes	93
5.2.1 Tamaño de la muestra	93
5.2.2 Intervenciones control	95
5.2.3 Enmascaramiento	96
5.2.4 Duración del estudio y seguimiento	97
5.2.5 Intervención experimental	99
5.2.6 Medidas de resultado	103
5.3 Las guías de reporte	107
5.3.1 ¿Son los elementos de las guías CONSORT suficientemente claros para los autores de manuscritos?	107
5.3.2 ¿Es posible que la utilidad de las guías CONSORT difiera dependiendo del campo de estudio?, ¿Es la extensión para intervenciones no-farmacológicas adecuada para todas las intervenciones no farmacológicas?	108
5.4 Fortalezas, limitaciones e implicaciones	109
5.4.1 Fortalezas	109
5.4.2 Limitaciones	110
5.4.3 Implicaciones	111
6. Conclusiones	114
7. Abreviaturas	118
8. Anexos	122
8.1 Anexo 1	124
8.2 Anexo 2	126
8.3 Anexo 3	130
8.4 Anexo 4	132
9. Bibliografía	142

1. INTRODUCCIÓN

1. Introducción

1.1 Fundamentos de la Terapia Manual

1.1.1 Definición

La Terapia Manual (TM) es un tratamiento físico utilizado para tratar pacientes con distintos problemas de salud, aunque su aplicación está especialmente extendida en el campo del dolor musculoesquelético. Consiste en la aplicación de fuerzas y movimientos a través de las manos con una finalidad terapéutica. Esta aproximación clínica incluye técnicas diagnósticas y de tratamiento para el tejido blando y las articulaciones con el propósito de modular el dolor, incrementar los rangos de movilidad, reducir y/o eliminar procesos inflamatorios y apoyar la reparación tisular. Su objetivo principal es facilitar el movimiento y mejorar la función biológica¹. Actualmente, la TM es utilizada por una gran variedad de profesionales y tiene un papel destacado en la atención sanitaria². Entre sus abordajes, la TM incluye técnicas dirigidas a los tejidos blandos, movilización/manipulación articular y neural y acupresión^{2,3}.

1.1.2 Historia de la Terapia Manual en Occidente

La primera descripción del uso de la TM se atribuye a Hipócrates (460-355 AC) en el documento *"On Setting Joints by Leverage"*. Hipócrates demostró la efectividad de varias técnicas manuales y propuso modificaciones de los parámetros de fuerza como la dirección, la velocidad y la frecuencia de las manipulaciones². Posteriormente, Galeno (129-201/216 DC) escribió y realizó numerosas ilustraciones de muchas de las técnicas manipulativas de Hipócrates¹. Él mismo aplicó también técnicas manuales en las extremidades y en la columna vertebral como el masaje y la manipulación espinal^{2,4}.

La edad media representó un vacío en el avance del conocimiento médico en el mundo occidental debido a la influencia que tuvo la Iglesia en la atención sanitaria. Es por ello que existe muy poca bibliografía del devenir de la TM en esa época¹. A partir del Renacimiento, en 1579, destaca la figura de Ambrose Pare, cirujano militar francés que utilizaba la manipulación en el tratamiento de las hipercifosis dorsales basándose en los trabajos de Hipócrates. Durante el siglo XVII y XVIII, la TM se desarrolla en Europa (especialmente en Inglaterra) con la aparición de los "bonesetters", terapeutas manuales que transmitían "su arte" de padres a hijos^{1,2}. Con el tiempo, esta modalidad terapéutica fue incorporándose a la práctica médica convencional, abandonando paulatinamente el contacto manual y enfatizando los abordajes farmacológicos. En Inglaterra, fue a mediados del siglo XX que los fisioterapeutas y los osteópatas heredaron la práctica de los bonesetters^{1,2}.

Es en los Estados Unidos (EEUU), a finales del siglo XIX, donde se encuentran los inicios de la Osteopatía y la Quiropráctica^{1,2,4}. Andrew Taylor Still (1828-1917), hijo de un practicante y pastor metodista, fundó la Osteopatía en 1874. Decepcionado por la práctica médica existente y devastado por la muerte de tres de sus hijos en una epidemia de meningitis, propuso un abordaje médico distinto con énfasis en la atención integral de la salud y el papel determinante del sistema musculoesquelético en ésta. Still propuso la interrelación entre la estructura y la función del cuerpo, y la capacidad innata del cuerpo para auto sanarse. Luego, propuso que esta interrelación podría mejorarse a través de técnicas ma-

nuales^{1,2,4,5}. En 1895, David Palmer (1845-1913) fundó la Quiropráctica. Palmer propuso que una vértebra puede subluxarse, comprometiendo estructuras en el agujero intervertebral y afectando las estructuras inervadas desde ese segmento, resultando en una enfermedad. Además propuso que el ajuste manual (manipulación) de la “subluxación” restablecería la salud de los tejidos afectados². En la jerga quiropráctica, esta filosofía terapéutica se conocería como “la regla del nervio”^{1,4}.

La Fisioterapia ya existía en aquella época en Europa bajo otros nombres (masaje, gimnasia médica, masaje y movimiento), pero la profesión se fundó formalmente en Inglaterra en 1899 y se estableció en 1921 en los EEUU². Las dos figuras médicas asociadas al desarrollo de la TM dentro de la Fisioterapia fueron James Mennell (1880-1957) y Edgar Cyriax (1874-1955)^{1,4}. James Mennell enseñó a los fisioterapeutas las técnicas de manipulación a principios del siglo XX en Inglaterra, al igual que hizo Cyriax a mediados del siglo XX. Tanto Mennell como Cyriax integraron la terapia manual en la medicina ortopédica, haciendo hincapié en la importancia del diagnóstico mecánico y el tratamiento conservador de los trastornos musculoesqueléticos². El trabajo de Mennell y Cyriax fue continuado por sus hijos John Mennell y James Cyriax⁴.

En los últimos 70 años, el desarrollo científico de la Fisioterapia ha sido exponencial y con ella, también ha mejorado el conocimiento científico de la TM. A partir de la década de los 50, los fisioterapeutas de todo el mundo comenzaron a investigar, a desarrollar la profesión y a organizarse⁴. Una de las figuras contemporáneas más relevantes en el campo de la TM fue el australiano Geoff Maitland (1924-2010). Maitland tuvo influencias de profesionales de la medicina física, osteópatas, quiroprácticos, bonesetters y especialmente de Mennell y Cyriax. Maitland fue pionero en introducir las técnicas oscilatorias, perfeccionó el concepto de barrera motriz e introdujo técnicas manuales diferentes a la manipulación con impulso (Thrust techniques)^{1,4}. Los principios de evaluación y tratamiento de Maitland han sido adoptados por terapeutas de todo el mundo⁶. El neozelandés Robin McKenzie (1931-2013) hizo también importantes contribuciones a la vez que argumentó sobre el exceso de importancia que tradicionalmente se le había dado a la TM dentro de la fisioterapia y desarrolló aproximaciones más activas. En Inglaterra, cabe destacar el papel de Gregory Grieve (1918-2001), considerado una de las personas más influyentes en el desarrollo de la fisioterapia moderna y un gran promotor de la visión científica de la TM. La colaboración entre Maitland y Grieve propició un gran progreso de los métodos educativos y clínicos utilizados en fisioterapia manipulativa en todo el mundo⁷. Otros fisioterapeutas destacados en el campo del progreso científico de la TM contemporánea fueron y siguen siendo el noruego Freddy Kaltenborn (1923-2019) y los neozelandeses Stanley Paris y Brian Mulligan (1958-2019).

1.1.3 Mecanismos de acción de la TM

A lo largo de la historia, los efectos de la TM han sido predominantemente explicados a partir de una concepción mecanicista del funcionamiento del cuerpo humano. Esta visión estructuralista y fuertemente basada en la biomecánica ha constituido, hasta muy recientemente, la base teórica de la TM. De hecho, sigue aún muy extendida entre los profesionales que la practican y los docentes^{8,9}. En el contexto de la aproximación estructuralista, se cree que las desviaciones posturales, las asimetrías corporales y la patomecánica son los factores predisponentes y de mantenimiento de muchas afecciones musculoesqueléticas. A su vez, se presenta la TM como una herramienta clínica para diagnosticar y tratar dichas disfunciones¹⁰. No obstante, los resultados de la investigación científica no solo han cuestionado estos modelos, sino que actualmente existen numerosas evidencias sobre la falta de plausibilidad y relevancia clínica de muchas de estas teorías¹⁰⁻¹⁴. Por esta razón, en las últimas décadas se han visto incrementadas las voces críticas y las llamadas a reconsiderar y actualizar el conocimiento existente sobre los mecanismos de acción de la TM^{9,10,13,15,16}.

La investigación realizada en el campo de las neurociencias y, muy específicamente, en el campo del dolor y la interocepción, han fomentado la aparición de modelos contemporáneos más acordes a la evidencia científica actual¹⁷⁻²⁰. Según los modelos actuales, los efectos de la TM se explican bajo una perspectiva neuromoduladora. Si bien se precisa un estímulo mecánico inicial, se reconoce que el efecto local conseguido es poco duradero y difícil de reproducir de forma fiable. Sin embargo, dicho estímulo es capaz de iniciar una serie de reacciones en cadena con potencial impacto en todas las divisiones del sistema nervioso (periférico, autonómico y central)^{17,18,20,21}. El efecto sobre las vías del dolor y los procesos de sensibilización (periférica y central) y el impacto en las vías interoceptivas explica, de una forma plausible, los efectos analgésicos registrados por la TM^{17-19,22-25}.

A pesar de que el efecto neuromodulador es actualmente el más aceptado, otros mecanismos fisiológicos se han descrito y relacionado con los efectos de la TM. La mecanotransducción, el mecanismo por el cual la perturbación mecánica influye en la expresión genética y el comportamiento celular, es un área de la biología molecular que experimenta un rápido desarrollo. Varios autores han señalado la relevancia de la mecanotransducción como mecanismos de acción de la TM²⁶⁻²⁹. Sin embargo, este tipo de investigación es muy reciente y muchas preguntas están todavía por responder en relación a la mecanotransducción. Queda mucho por descubrir acerca de su relevancia en la salud, la enfermedad y las intervenciones terapéuticas como la TM. Por ejemplo, no se conoce la dosificación ideal de las cargas mecánicas, ni cuándo es mejor utilizarlas durante el proceso de reparación para producir los efectos óptimos^{26,27}.

Finalmente, en el campo de las terapias físicas, y muy particularmente de la TM, los efectos inespecíficos desempeñan un papel clínico muy relevante. Los mecanismos vinculados a la respuesta placebo en el contacto terapéutico manual han sido ampliamente documentados³⁰⁻³⁴. Esta respuesta se atribuye a elementos contextuales vinculados con el ritual terapéutico que acompaña dichas intervenciones. La relación terapeuta-paciente, el entorno clínico así como las expectativas vinculadas al tratamiento son factores muy determinantes en el resultado final³⁵⁻³⁸.

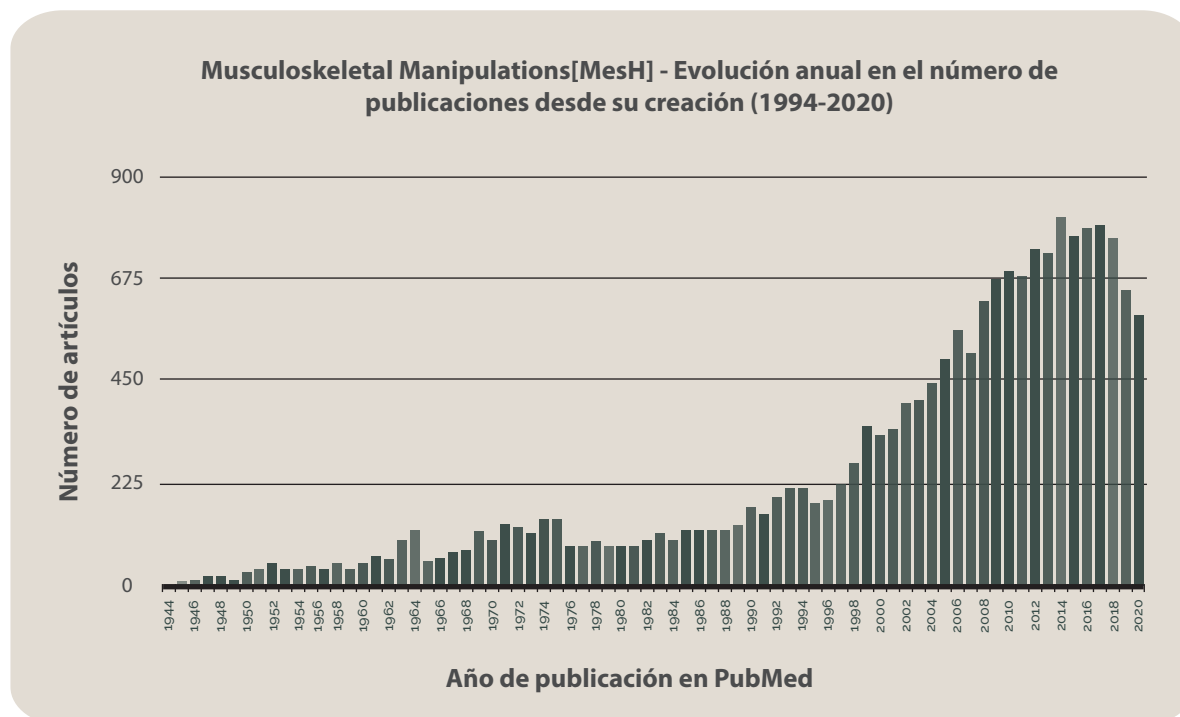
1.1.4 Literatura científica en TM

En paralelo al aumento de publicaciones registrado en el campo de la fisioterapia en general^{39,40}, la producción científica en el ámbito específico de la TM ha aumentado exponencialmente en las dos últimas décadas. Este incremento se ve reflejado tanto en el número de artículos publicados como en la aparición de nuevas revistas y bases de datos bibliográficas (**Fig 1**). De hecho, la base de datos *PEDro* (*Physiotherapy Evidence Database*) - base de datos bibliográfica que contiene ensayos aleatorizados, guías de práctica clínica y revisiones sistemáticas en el campo de la fisioterapia - inició su actividad en Octubre del año 2000.

Sin embargo, la publicación de la investigación realizada por la fisioterapia moderna puede ser considerada un fenómeno relativamente nuevo en la literatura científica. La revista más citada en este campo, *Physical Therapy*⁴¹ no se incluyó en la lista ISI (*Institute for Scientific Information*) hasta el año 2005 y no recibió una puntuación de factor de impacto hasta el 2007. Además, hay que tener en cuenta que, aproximadamente, menos de la mitad de los estudios que se realizan en el campo de las medicinas complementarias llegan a ser publicados (en revistas con factor de impacto) y que no todas las revistas específicas de fisioterapia o de TM tienen factor de impacto⁴². Este dato es especialmente relevante en el caso de los estudios publicados en español⁴³. Actualmente, la base de datos que ofrece la mayor tasa de cobertura de los títulos más citados en fisioterapia es CINAHL (*Cumulative Index to Nursing and Allied Health Literature*), seguido de MEDLINE (*Medical Literature Analysis and Retrieval System Online*) y EMBASE (*Excerpta Medica dataBASE*)⁴¹.

La TM está bien establecida como campo multidisciplinario y esto se refleja tanto en el perfil de los lectores de revistas como en la autoría de los artículos de TM⁴². A modo de ejemplo, en el journal *Manual Therapy* (actualmente llamado *Musculoskeletal Science and Practice* - ISSN: 2468-7812), si bien la mayoría de los lectores y autores son fisioterapeutas que trabajan en el campo de la terapia manual, la revista también ha publicado trabajos escritos por médicos, investigadores de otros campos, quiroprácticos y osteópatas⁴⁴. De la misma manera, la producción científica en el campo de la rehabilitación trasciende las revistas y bases de datos especializadas en el campo de la fisioterapia⁴⁵ y por extensión, lo mismo ocurre con la literatura sobre TM. La *National Library of Medicine* indexa el término "Terapia Manual" bajo el MESH (Medical Subject Heading) "*Musculoskeletal Manipulations*" ("*Various manipulations of body tissues, muscles and bones by hands or equipment to improve health and circulation, relieve fatigue, promote healing*"). Este MESH se encuentra ubicado tanto bajo la categoría "*Physical Therapy Modalities*" como bajo la categoría "*Complementary Therapies*" ejemplificando de nuevo la amplitud del término y la heterogeneidad de las referencias potencialmente indexadas en él.

Fig 1: Musculoskeletal Manipulations[MesH] - Evolución anual en el número de publicaciones desde su creación (1994-2020)



1.2 Investigación en Terapia Manual

1.2.1 El Ensayo Clínico Aleatorizado en Terapia Manual

El primer ECA realizado y publicado del que se tiene constancia en la literatura fue realizado en 1948 para evaluar el efecto de la estreptomycin en el tratamiento de la tuberculosis pulmonar. A partir de aquí, este método se convirtió en una herramienta imprescindible en la revolución farmacológica posterior a la segunda guerra mundial. Históricamente, el ECA ha sido utilizado en el contexto de la investigación biomédica y es una herramienta indispensable previa a la autorización de nuevos fármacos a partir de 1962⁴⁶. Desde entonces hasta nuestros días, el ECA ha sido considerado el *gold standard* para la evaluación de intervenciones terapéuticas. Ello se debe a que las características de su diseño minimizan los potenciales sesgos introducidos tanto por los investigadores como por los pacientes confiriéndole una gran validez interna^{47,48}. Ligado a su alta capacidad para reducir sesgos, en la realización de un ECA se asumen una serie de premisas universales únicas para este diseño (**Tabla 1**). Sin embargo, hay que tener en cuenta que, incluso en las situaciones más controladas, a menudo varias de estas premisas pueden no cumplirse⁴⁹.

Tabla 1: Premisas asumidas en la realización de un ECA. (Adaptado de Walach et al. 2006)

Equilibrio	<i>El paciente y el terapeuta no tienen preferencia por un tratamiento</i>
Falta de conocimiento previo	<i>Realmente se desconoce cuál de las dos alternativas es “mejor” y no hay pruebas suficientes sobre los efectos del tratamiento de otras fuentes.</i>
Preferencia por la especificidad	<i>Solo los efectos específicos atribuibles a la intervención son terapéuticamente válidos.</i>
Independencia del contexto	<i>Existe una magnitud “verdadera” de eficacia, o un tamaño del efecto estable independiente del contexto</i>
Validez externa y ecológica conocida	<i>El conocimiento sobre un efecto terapéutico extraído de un ECA es fácilmente transferible a la práctica clínica, si los criterios de exclusión e inclusión del ensayo coinciden con las características de un paciente determinado.</i>

Por otro lado, existen muchas prácticas en la atención sanitaria que no siguen la lógica de una intervención farmacológica. La enfermería, la psicología, la cirugía o la rehabilitación son solo algunos ejemplos. Los estudios de investigación en rehabilitación suelen comparar diferentes paquetes terapéuticos y no intervenciones específicas, aunque a menudo se basan en un elemento particular dentro de cada paquete como puede ser la TM. En estas condiciones, el control total de las interacciones entre las intervenciones y su contexto (una de las premisas fundamentales del ECA) es difícil de lograr⁵⁰. De hecho, en la evaluación de terapias no-farmacológicas y, muy en particular en las llamadas Intervenciones Complejas (ICs), puede no darse ninguna de las premisas que generalmente se asumen para los estudios farmacológicos⁴⁹. Además, al intentar diseñar un ECA ordinario de alta calidad para la evaluación de una IC topamos con un sinfín de dificultades metodológicas y aparece el reto mayúsculo de conseguir un equilibrio aceptable entre la validez interna y la validez externa del estudio. En el campo de la rehabilitación y la TM, estas dificultades y limitaciones han sido ampliamente reportadas y discutidas por fisioterapeutas, osteópatas y quiroprácticos, pero también por médicos e investigadores^{31,50-61}. A continuación, se resumen los puntos más destacados.

- I. **Homogeneidad en los grupos de estudio:** Uno de los requisitos que se pretende conseguir con el diseño de un ECA es la homogeneidad de los grupos de estudio. Para tal propósito, a menudo se establecen estrictos criterios de inclusión y exclusión que permitan obtener muestras comparables. Sin embargo, la extrema rigidez en estos criterios puede conllevar que se incluya solo una pequeña proporción de aquellos pacientes que se ven rutinariamente en la práctica clínica que, en este campo concreto, son especialmente heterogéneos^{59,61}. Además, la excesiva homogeneidad de los grupos puede dificultar la detección del efecto de la intervención de estudio en un sub-grupo diferenciado de otro. En el caso de existir un efecto positivo en un determinado subgrupo, este quedaría diluido por la falta de

efecto en otros/s grupo/s³¹. Además, la hiperselección de los sujetos puede comprometer la variabilidad en los factores psicosociales de la muestra, siendo estos factores determinantes en la evaluación del efecto de la TM^{13,62}.

- II. **El grupo control:** La elección de la intervención de control es el primer paso y posiblemente el más importante en el diseño de un ECA. La elección de una intervención de control debe estar guiada por el principio ético de equilibrio (*equipoise*) o “principio de incertidumbre”. Este principio sostiene que los pacientes solo deben ser enrolados en un ECA si existe una incertidumbre sustancial sobre cuál de los tratamientos del ensayo podría ser más beneficioso⁶³. Los estudios en los que se cree que los grupos de intervención y control no son equivalentes violan el principio de incertidumbre y pueden dar lugar a resultados sesgados^{63–65}. En los ensayos farmacológicos, la administración de placebos inertes en el grupo control facilita las comparaciones cegadas. Sin embargo, la elección de controles creíbles que minimicen el sesgo de detección en los ECA en TM plantea desafíos únicos para los investigadores⁵⁰. A pesar de los esfuerzos realizados para encontrar un control placebo (o simulación “sham”) adecuado para la investigación en TM^{66–70}, todavía no existe una guía o un modelo estandarizado y validado para tal fin⁷¹. Existen al menos dos problemas importantes al diseñar un estudio que utiliza controles que no reciben intervención manual. Primero, los sujetos pueden deducir que han sido asignados aleatoriamente al grupo de control (es decir, desenmascarados), lo que los hace más propensos a abandonar el ensayo y/o a responder negativamente a la intervención fruto del sesgo de detección (respuesta nocebo). En segundo lugar, los sujetos de control sin intervención o intervención no-manual pueden recibir menos atención clínica (ya sea real o percibida), lo que introduce un problema ético^{50,56}. Por otro lado, varios estudios han demostrado que el tacto por sí solo podría contribuir significativamente al efecto placebo en intervenciones manuales^{68,72–74}. En este escenario, los investigadores se ven comúnmente forzados a utilizar controles activos o controles basados en el “*standard care*” que pueden no ser óptimos para contestar su pregunta de investigación. Ello también puede conllevar potenciales problemas en el reclutamiento y modificaciones en las expectativas del paciente hacia el tratamiento, especialmente en el periodo post-aleatorización. Si el tratamiento finalmente asignado no coincide con las creencias y expectativas de los pacientes, es probable que el efecto del tratamiento sea moderado⁶¹. Es lo que se conoce como efecto Hawthorn⁷⁵.
- III. **Enmascaramiento:** Íntimamente ligado a la dificultad para seleccionar un control adecuado a las intervenciones de TM, existen las limitaciones relacionadas con el enmascaramiento de la intervención. De hecho, al comparar ECAs que evalúan tratamientos no-farmacológicos versus farmacológicos, la falta de cegamiento en los artículos sobre tratamientos no-farmacológicos explica la mayoría de las diferencias metodológicas⁷⁶. Puesto que se trata de una intervención directamente realizada por el terapeuta, es prácticamente imposible que este pueda estar enmascarado en relación a los grupos de estudio por lo que tendrá conocimiento real del tratamiento que aplica^{31,56}. Debido a las dificultades mencionadas anteriormente en la selección de un buen control, el enmascaramiento de la intervención a los pacientes en los ECA de TM constituye también un reto para los investigadores. Si

no se puede aplicar un placebo/sham confiable, va a ser muy difícil cegar a los pacientes. La falta de enmascaramiento de la intervención a los pacientes conduce a distintos problemas metodológicos incluyendo potenciales sesgos de realización, aumento de la tasa de abandonos o adherencia al protocolo de investigación y alteraciones de la estimación del efecto debido a una distribución desigual de las expectativas de los pacientes^{31,54}. No obstante, si bien el enmascaramiento de la intervención a terapeutas y pacientes presenta múltiples dificultades en este campo, el enmascaramiento respecto al evaluador de los resultados es mucho más sencillo de conseguir. En estos casos es importante el uso de evaluadores externos para asegurar el cegamiento y se recomienda también extender el cegamiento al responsable del manejo y análisis estadístico de los datos si es que se trata de personas diferentes⁵⁰. En los ECA publicados en los últimos años, el enmascaramiento del evaluador parece haberse convertido en un método más común³¹.

- IV. **Estandarización de las intervenciones:** Una característica metodológica importante es cualquier diseño experimental (y especialmente en los ECA) es la estandarización de la intervención. Esto se considera importante para controlar los posibles sesgos de realización que comprometen la validez interna del estudio y para asegurar la reproducibilidad del ECA. Sin embargo, la estandarización de la intervención choca frontalmente con la práctica de la TM clínica tradicional donde las intervenciones se adaptan a las necesidades, evolución y características del paciente. De nuevo, emerge el conflicto entre la validez interna y externa⁵⁵. Las ICs dependen en gran medida del contexto, y su aplicación requiere una adaptación significativa a éste para que su práctica sea eficaz y su evaluación realista. Sin embargo, con demasiada frecuencia, una IC se reduce a sus partes constitutivas para que cumpla los estrictos requisitos de un ECA^{77,78}. Es por ello que los investigadores que pretenden realizar un ECA en TM deben encontrar la manera de reproducir la práctica clínica a la vez que cumplen con el rigor metodológico esperado de un diseño experimental de alta calidad. Si bien la evaluación de técnicas concretas puede estar ceñida a un protocolo estandarizado, la evaluación de programas de tratamiento con distintas sesiones suelen estar sujetos a cierto grado de individualización.
- V. **Efectos contextuales:** Otro de los objetivos fundamentales de los ECA consiste en ofrecer una valoración “neta” del efecto de la intervención de estudio en una población concreta (**Tabla 1**). Es decir, se pretende aislar el “ingrediente activo” del resto de factores contextuales que envuelven la intervención para evaluar su efecto “objetivo”⁵⁴. Si bien algunos efectos contextuales pueden ser controlados mediante los procedimientos de aleatorización, aislar todos los efectos específicos de un tratamiento de aquellos no específicos es, de nuevo, un desafío para los investigadores interesados en determinar la eficacia de la TM^{50,56}. En el caso del dolor musculoesquelético y su tratamiento, se han descrito multitud de factores contextuales presentes en distintas dimensiones del encuentro terapéutico³⁵ y estos impactan también en el contexto investigador⁶¹. Varios estudios han evidenciado la gran importancia clínica de estos efectos contextuales. De hecho, una reciente revisión realizada para evaluar el impacto de dichos efectos en los ECAs controlados con placebo, concluye que aproximadamente la mitad del efecto general del

tratamiento parece atribuible a efectos contextuales más que al efecto específico de las intervenciones⁷⁹. Por ejemplo, en ECAs realizados para la evaluación de tratamientos para la artrosis, se concluye que el 75% del efecto general se puede atribuir a efectos contextuales más que al efecto específico de los tratamientos⁸⁰. También, en ECAs sobre intervenciones para la fibromialgia, se estima que más de la mitad del efecto del tratamiento se debe a factores contextuales no específicos⁸¹. En el campo concreto de la TM, las creencias y expectativas de un individuo sobre la eficacia del tratamiento pueden tener un efecto muy significativo en los resultados comúnmente medidos, incluido el dolor, la función y la calidad de vida^{34,50}. De hecho, varios investigadores en este campo abogan por utilizar deliberadamente el potencial terapéutico de estos efectos contextuales considerándolos parte esencial e indisoluble de la intervención^{82,83}. En conclusión, parece que una de las bases fundamentales en el diseño de los ECAs, al menos en su versión más explicativa, es de nuevo complicado de conseguir en las intervenciones manuales. Pero no solo eso, sino que atendiendo a la filosofía terapéutica de algunas de las TM, incluso puede considerarse no pertinente⁵⁷.

VI. **Tamaño de la muestra:** El tamaño de la muestra es también un aspecto fundamental en los ECAs. Los cálculos del tamaño de la muestra se centran en las pruebas de hipótesis estadísticas y, evidentemente, es deseable que la potencia sea lo más alta posible. Sin embargo, para aumentar la potencia, el tamaño de muestra requerido también debe aumentar. Es fundamental que los investigadores calculen el tamaño de muestra óptimo⁸⁴. Si el tamaño de la muestra es demasiado pequeño, el ensayo no podrá demostrar el (menor) efecto de interés clínico, en el caso de que este existiera en la población de estudio^{84,85}. Sin embargo, los estudios en TM se realizan comúnmente utilizando muestras pequeñas³¹. Al no ser una práctica estandarizada y normalizada en todos los sistemas de salud públicos e institucionales, gran parte de la investigación se realiza en consultas privadas y/o universidades, desfavoreciendo la realización de grandes ensayos multicéntricos y conllevando dificultades en el reclutamiento de sujetos. El problema de las muestras pequeñas es de especial relevancia, pues estudios en TM han evidenciado mejores resultados cuando la muestra puede ser dividida o agrupada en sub-grupos⁸⁶. Finalmente, otro problema con los tamaños de muestra pequeños es que la comparabilidad de los grupos de estudio al inicio (baseline) puede estar comprometida. Solo con un número adecuado de pacientes aleatorizados podemos tener alguna seguridad de que los factores pronósticos (conocidos y desconocidos) se distribuirán de manera uniforme entre los grupos de estudio³¹.

VII. **Generabilidad:** Todos los puntos anteriormente comentados evidencian el conflicto existente entre los requerimientos metodológicos exigidos en el diseño de un ECA y la realidad de la práctica clínica en TM. De hecho, los ECA tienen una validez externa a menudo limitada porque están diseñados para proporcionar una sólida evidencia de que el tratamiento está teniendo realmente un efecto en la población de estudio⁸⁷. Sin embargo, los mecanismos que aseguran la validez interna de un ECA (por ejemplo, la asignación al azar y la estandarización de la intervención), amenazan la generalización y, por lo tanto, la relevancia de los resultados⁵⁵. Por lo tanto, puede considerarse que la falta de generabilidad es la

debilidad principal de este tipo de estudio al haber varios elementos del diseño y del reporte de los ECAs que limitan su validez externa^{88,89}. Si bien este conflicto está siempre presente en mayor o menor grado en la evaluación de cualquier intervención sanitaria, en el caso de las ICs, las intervenciones no-farmacológicas y la TM en particular, es especialmente remarcable. De hecho, en TM, los problemas asociados con el uso de ECAs están menos relacionados con la validez interna del diseño del estudio y más con la validez externa de los resultados⁵⁵. La investigación en TM es fundamentalmente llevada a cabo por profesionales con un perfil mixto clínico-investigador, siendo el perfil exclusivamente investigador muy minoritario. En el momento de realizar un ECA, estos investigadores deben decidir los parámetros de un diseño que en muchos aspectos “no encaja” con la manera en que se practica la TM y eso deriva en estudios considerados o de poca calidad o con importantes limitaciones para la generabilidad. Sin embargo, intentar encontrar una solución al conflicto entre validez interna y validez externa es un ejercicio infructuoso. La cuestión es simple: si mejora la calidad científica del método (validez interna) se reduce la generabilidad (validez externa), y viceversa⁵¹. Es importante remarcar también que, para emitir juicios sobre la validez externa de un ECA, siempre se requiere de conocimiento experto de la materia⁸⁷. Para algunos autores, el problema reside básicamente en la necesidad de encontrar la “evidencia adicional” que se precisa para aplicar los resultados de un determinado ECA a otras poblaciones⁹⁰. No obstante, si los resultados de un ECA no pueden ser fácilmente generalizables a un paciente definido en un entorno clínico particular y con una práctica clínica particular, entonces la utilidad del ECA debe ser cuestionada en términos de su capacidad para informar la práctica en TM⁵⁵.

Los siete puntos descritos anteriormente resumen la mayoría de las limitaciones y dificultades a las que los investigadores en TM se enfrentan en el momento de realizar un ECA. Se concluye de todo ello que, para valorar el efecto específico de técnicas aisladas, el ECA representa probablemente la mejor opción. Sin embargo, para mejorar la investigación de tratamientos manuales, deben explorarse adaptaciones metodológicas al ECA que permitan un mayor equilibrio entre validez interna y externa, se debe fomentar la búsqueda y aplicación de métodos alternativos y no menospreciar diseños considerados a priori de peor calidad pero que se ajustan mejor a este entorno clínico particular^{49,51,54,55,57,59,61}.

1.2.2 La Terapia Manual como una intervención compleja

Una intervención compleja (IC) se define como aquella compuesta de varios componentes que interactúan entre sí. Este tipo de intervenciones presentan una serie de particularidades especiales para los investigadores. El principal problema se relaciona con la dificultad de estandarizar el diseño y la ejecución de estas intervenciones, la dificultad práctica de aplicar métodos de investigación experimentales y la complejidad de los factores causales que explican la asociación entre la intervención y sus resultados^{91,92}. En el año 2000, el **Medical Research Council** (MRC) publicó el primer marco teórico (*framework*) sobre cómo deben desarrollarse y evaluarse las IC⁹³, una guía que ha sido posteriormente actualizada en el año 2008⁹² y luego en el 2019⁹¹. Sin embargo, existen distintos niveles de complejidad según la intervención (**Tabla 2**).

Tabla 2: Componentes de una intervención compleja. (*Adaptado de Craig et al 2019*)

Dimensiones de complejidad
● Número e interacciones entre componentes dentro de las intervenciones experimentales y de control ● Número y dificultad de los comportamientos requeridos por quienes realizan o reciben la intervención ● Número de grupos o niveles organizacionales a los que se dirige la intervención ● Número y variabilidad en los resultados ● Grado de flexibilidad o adaptación que la intervención permitió
Implicaciones para el desarrollo y la evaluación
● Se necesita una buena comprensión teórica de cómo la intervención provoca el cambio, de modo que se puedan identificar y fortalecer los puntos débiles de la cadena causal. ● La falta de impacto puede reflejar una falla en la implementación (o problemas iniciales) más que una ineficacia genuina; Se necesita una evaluación exhaustiva del proceso para identificar problemas de implementación. ● La variabilidad en los resultados a nivel individual puede reflejar procesos de nivel superior; Es posible que los tamaños de las muestras deban ser más grandes para tener en cuenta la variabilidad adicional, y se deben considerar diseños aleatorizados por conglomerados (clusters) en lugar de individualmente ● La identificación de un único resultado primario puede no conllevar un mejor uso de los datos; Se necesitarán una serie de medidas y, cuando sea posible, se recogerán los efectos adversos. ● Asegurar una estricta fidelidad a un protocolo puede ser inapropiado; la intervención puede funcionar mejor si se permite la adaptación al entorno local.

Los procesos de rehabilitación, dentro de los cuáles se engloban las terapias físicas y en particular la TM, responden a la denominación de IC por estar comúnmente compuestos por diversas técnicas que se combinan para lograr el objetivo clínico deseado⁵⁰. Estas intervenciones se configuran en “paquetes terapéuticos” que pueden incluir varios abor-

dajes y aproximaciones diferentes⁵⁰. De hecho, se considera que los estudios de intervenciones en fisioterapia que utilizan protocolos de intervención altamente estandarizados, aunque metodológicamente rigurosos, pueden no reflejar necesariamente la práctica clínica del “mundo real” y, en muchos casos, los resultados pueden ser decepcionantes⁹⁴. Si lo centramos en el campo de la TM, un tratamiento convencional incluye distintas técnicas empleadas en distintas sesiones que, a su vez, pueden variar entre sí. Además, los resultados obtenidos son múltiples y abarcan distintas dimensiones del paciente (percepción de dolor, rango de movilidad, recuperación de la autonomía, etc.). Finalmente, el tratamiento es individualizado y debe adaptarse a las características y forma de ser de cada paciente. Todos estos elementos, comunes en la práctica clínica, convierten a la TM en una IC y dicha complejidad conlleva dificultades para entender y evaluar el mecanismo de acción que hay detrás de su efecto⁵⁴.

1.2.3 Características de la investigación en Intervenciones complejas

En el contexto de la ICs, los diseños experimentales formales y de corte más explicativo son de difícil aplicación. El desarrollo de una intervención compleja que incluye múltiples modalidades de tratamiento y la selección de técnicas individualizadas requiere un enfoque sistemático para diseñar la intervención teniendo en cuenta aspectos relativos a su marco teórico, a la evidencia disponible y a las limitaciones prácticas⁹⁴. No existe un método ideal para la evaluación de una IC sino que se debe utilizar una variedad de métodos enfocados a abordar distintas dimensiones de la intervención, tales como la evaluación del efecto de la intervención, la evaluación de su implementación, los mecanismos de acción y el efecto de los factores contextuales⁹⁵. Tanto es así, que la evaluación de la calidad de la evidencia a través de la visión piramidal tradicional establecida puede carecer de sentido⁵⁵, habiéndose propuesto modelos alternativos más acordes con las características de las ICs^{49,95}. En definitiva, tal y como se apunta en el framework del MRC⁹¹, la evaluación de las ICs consiste principalmente en un enfoque global que comprende el desarrollo de una intervención, desde su marco teórico hasta los varios pasos de su evaluación⁹⁶.

Por todo ello, previo a la elección del método de investigación, es necesario preguntarse qué elementos y dimensiones de la IC queremos explorar y/o evaluar. En otras palabras, el desafío consiste en identificar el diseño que encaje mejor con el objeto de evaluación, teniendo en cuenta que estos son múltiples y que van más allá de la evaluación del efecto o la eficacia⁹⁵. El modelo circular propuesto por *Walach et al.*⁴⁹ derivado de las ciencias sociales reconoce que existen métodos óptimos para responder preguntas específicas y que una combinación de todos los métodos constituye la mejor evidencia científica (**Fig 2**). Los métodos experimentales que prueban específicamente la eficacia (ensayos clínicos aleatorizados – ECAs) deben complementarse con métodos de observación no experimentales que, por su naturaleza descriptiva, describan efectos y su aplicabilidad en la vida real. *Walach y cols.* proponen hasta ocho diseños distintos para evaluar cuatro dimensiones principales: i) eficacia (ECAs controlados con placebo y ECAs comparados con lista de espera), ii) efectividad comparada (ECAs con controles activos y estudios de controles apareados), iii) efectividad (estudios de cohortes y auditorías) y iv) efectos/seguridad (estudios observacionales y estudios de supervivencia)⁴⁹. Más recientemente, *Minary y cols.* han propuesto un framework metodológico enfocado al diseño y la evaluación de ICs⁹⁵ a partir de una revisión sistemática de estudios donde se han evaluado varias de estas in-

tervenciones. Dicha revisión sistemática incluyó un total de 105 estudios de los cuales 29 eran estudios de prevención o de promoción de la salud y 76 estudios clínicos o de investigación en servicios de salud. Los autores coinciden en la idea de que varios métodos son necesarios y deben ser usados sucesivamente y de forma combinada en varias etapas del proceso evaluativo. El modelo de *Minary y cols.* describe tres ejes (**Fig 3**). En el eje x encontramos el continuo entre validez interna y validez externa. En el eje y una vertical que marca el continuo entre el proceso, los mecanismos de acción de la intervención y sus efectos. Finalmente, un eje transversal posiciona los distintos diseños en función de la pregunta de investigación, transicionando desde los estudios de eficacia a los de implementación de la investigación. Dentro de este marco y situados en distintos cuadrantes, encontramos: adaptaciones al ECA (ECAs individuales, ECAs en conglomerado o ECAs pragmáticos) y métodos alternativos al ECA (estudios cuasi-experimentales, estudios de cohortes, evaluaciones realistas, estudios de casos y otros)⁹⁵.

Fig 2: Modelo circular de *Walach et al. 2006* para la evaluación de las intervenciones complejas. (Figura Incluida con permiso de los autores)

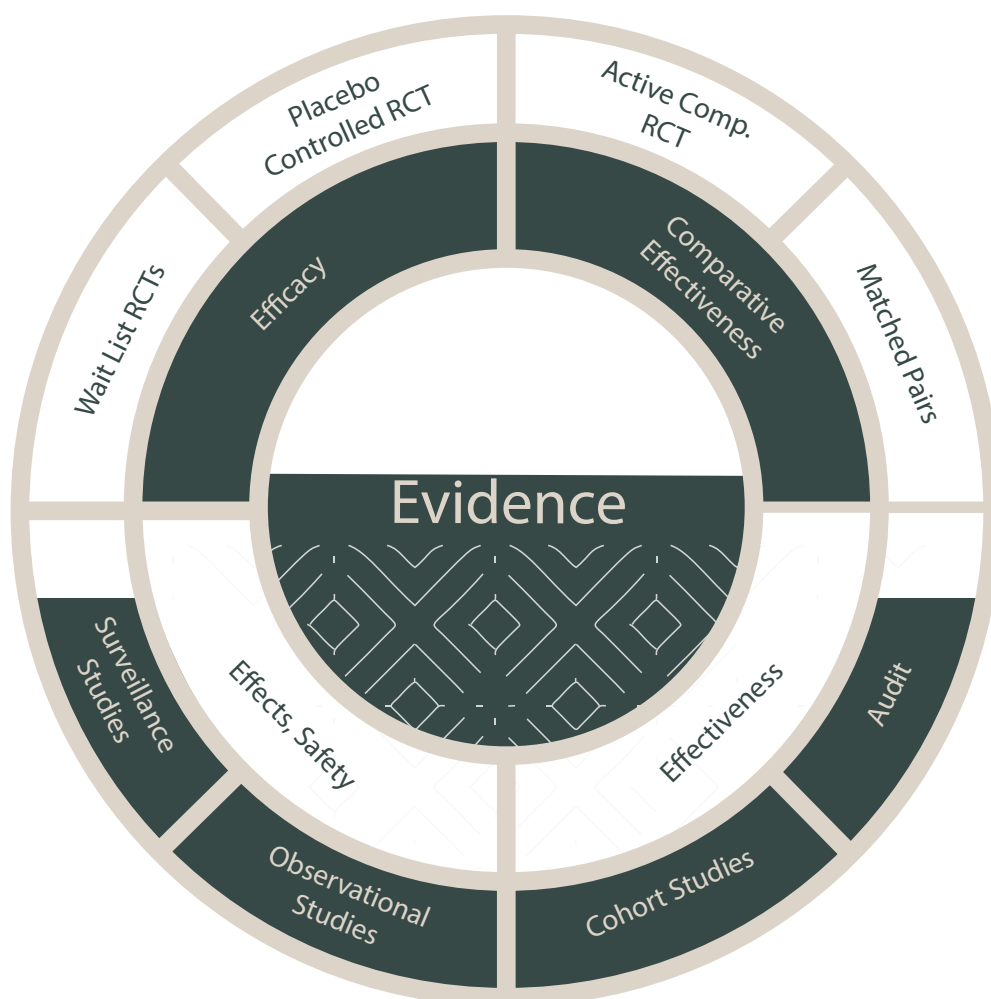
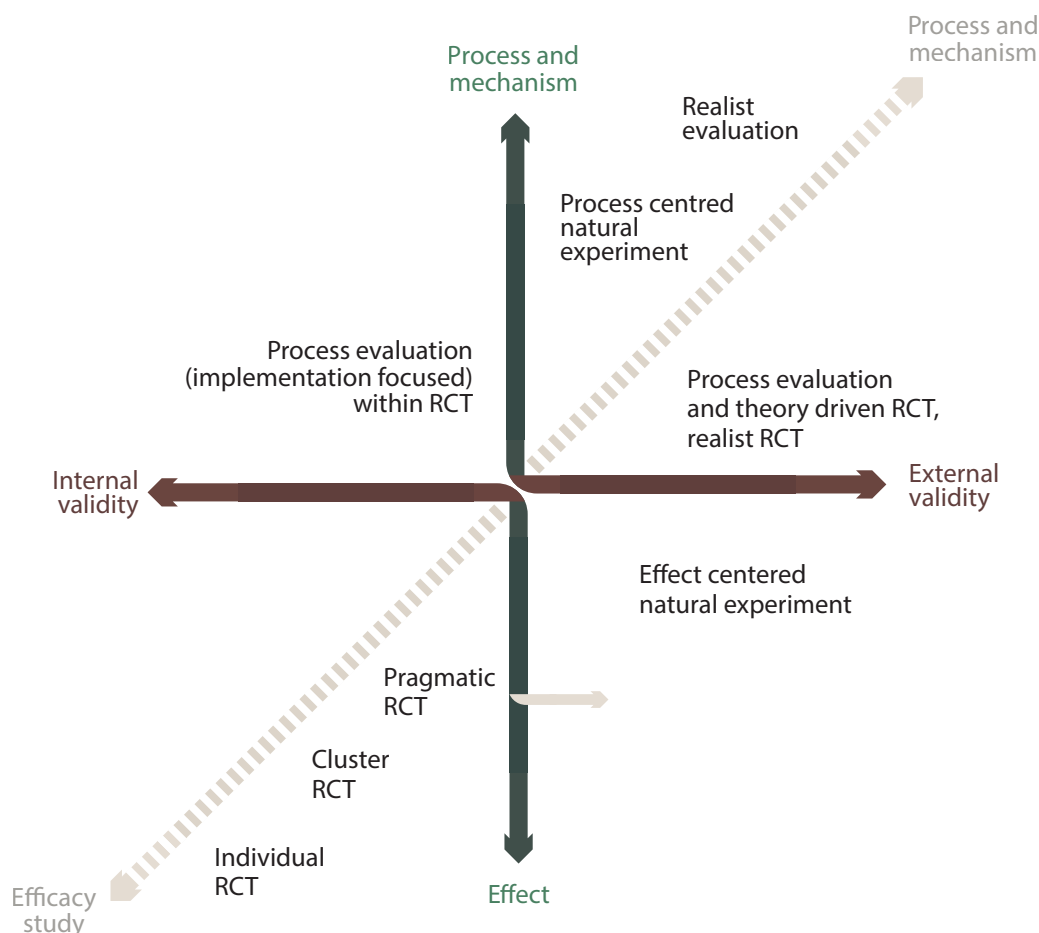


Fig 3: Framework de Minary et al. 2019 situando los diseños en relación a las preguntas de investigación. (Figura Incluida con permiso de los autores)



En conclusión, la combinación de métodos de investigación es necesaria para la correcta evaluación de las ICs. Sin embargo, tal y como nos muestra la literatura descrita anteriormente, si analizamos intervenciones terapéuticas, en algún punto de la evaluación es imprescindible recurrir al ECA o a alguna adaptación del ECA para establecer los efectos de dicha intervención. Entonces, la pregunta es: ¿Podemos utilizar el clásico diseño de ECA para la evaluación de una IC como la TM? La respuesta es sí, pero con dificultades^{54,77,78}. De hecho, aunque es fácilmente transferible a cualquier otro diseño, el framework propuesto por el MRC se desarrolló poniendo especial énfasis en aquellas intervenciones evaluadas a través de un ECA. En la guía específica publicada a tal efecto⁹⁷ se propone una evaluación secuencial en cinco etapas que incluyen una fase pre-clínica (o teórica) y cuatro fases clínicas (I-Modelo, II-Ensayo exploratorio, III-Ensayo Definitivo y IV-Implementación a largo plazo (**Fig 4**). Sin embargo, a pesar de que se han publicado buenos trabajos en el campo de la Fisioterapia basados en este modelo^{94,98}, seguir de forma estricta esta secuencia introduce muchas dificultades a los investigadores en este campo que, a menudo, cuentan con pocos o nulos recursos e investigan en entornos muy variados⁵⁴. Además, la evolución de las intervenciones en el cuidado de la salud y su posterior evaluación no siempre transcurre en etapas claramente definidas⁷⁸. Si bien el modelo propuesto por el MRC constituye el proceso teóricamente óptimo para evaluar una IC a través de un ECA, diferentes alternativas y adaptaciones al ECA pueden y deben ser contempladas.

Fig 4: Fases secuenciales del desarrollo de ECAs de intervenciones complejas.
(Adaptado de Campbell et al. 2000)



1.2.4 Calidad de la evidencia en Fisioterapia y Terapia Manual

El análisis de la literatura científica de los últimos 20 años nos informa que, a pesar del aumento en la producción científica experimentado en el campo de la TM, hay una importante falta de evidencia de alta calidad acerca de su efectividad^{31,40,99-102}. La baja calidad en los ECAs impacta directamente en su capacidad para informar la práctica clínica más allá de su limitada validez externa. Por ejemplo, los artículos que son de baja calidad metodológica (o tienen un alto riesgo de sesgo) tienden a sobreestimar los efectos del tratamiento¹⁰³. Por ejemplo, en el campo del dolor lumbar, se ha evidenciado que estudios con mayor riesgo de sesgo informan de tamaños del efecto que fueron, en promedio, 50% mayores que las estimaciones realizadas en los ECA con un menor riesgo de sesgo¹⁰⁴. Aunque la valoración de la calidad de la evidencia también está sujeta a elementos externos al propio estudio evaluado¹⁰⁵ y a que dicha valoración puede variar según la escala que se use¹⁰⁶, entre las varias posibles razones que explican la baja calidad de los ECA en fisioterapia y TM se encuentran las dificultades metodológicas descritas en la sección anterior. Prueba de ello es que los principales elementos identificados como responsables de esta falta de calidad se relacionan de una manera u otra con las dificultades anteriormente comentadas^{40,53,100,102}. Otro elemento determinante de la baja calidad de los ECA en este campo es la pobre adhesión que muestran los estudios a las guías de reporte estandarizadas. De nada sirve diseñar un estudio con un gran rigor metodológico si los parámetros concretos del diseño y de la conducción del experimento no se reportan adecuadamente

en el manuscrito final. Una revisión de ECAs en fisioterapia para problemas musculoesqueléticos¹⁰² y otra en el campo específico de la TM¹⁰⁰ revelan una serie de predictores que se asocian a una mejor calidad metodológica medida a través de la escala PEDro. Según estos dos estudios, uno de los predictores determinantes que se asocia a la calidad de los ECA en fisioterapia hace referencia al buen reporte de los mismos y que estos estén publicados en revistas que hayan manifestado su adhesión a las guías de reporte estandarizadas. En el caso de los ECAs, la guía CONSORT (*Consolidated Standards for Reporting Trials*).

1.2.5 CONSORT y el reporte en intervenciones no-farmacológicas

Con el objetivo de mejorar significativamente la calidad del reporte de los ECAs, surge en 1996 la Declaración CONSORT¹⁰⁷. Desde su primera aparición, este documento ha sido actualizado hasta su última versión publicada en 2010¹⁰⁸. La guía CONSORT está formada por unas recomendaciones destinadas a orientar a los autores sobre la información mínima necesaria a incluir en el informe publicado de un ECA. Ofrece, por lo tanto, un modelo estándar para que los autores preparen informes de los resultados de los ensayos, lo que facilita su informe completo y transparente y ayuda a su evaluación e interpretación críticas (<http://www.consort-statement.org>). Las principales y más destacadas revistas médicas (tanto generales como especializadas) y organizaciones editoriales líderes han manifestado su adhesión a la declaración CONSORT.

En 2014, 28 revistas especializadas en el campo de la rehabilitación acordaron adoptar una postura más decidida hacia la implementación de las guías de reporte con el fin de mejorar la calidad de los informes científicos en este campo¹⁰⁹. Sin embargo, a pesar de la mejora gradual a lo largo del tiempo¹¹⁰, sigue siendo necesario mejorar tanto la calidad metodológica de los ECAs como su reporte¹⁰². Centrándonos específicamente en la literatura sobre TM, existe evidencia que sugiere que la calidad del reporte no ha mejorado suficientemente en los últimos años¹⁰⁰, lo que, a su vez, sugeriría que el uso actual de las guías CONSORT no es óptimo. Sin embargo, es importante remarcar que la misma evidencia afirma que estas guías incluyen varios ítems con definiciones poco claras, lo que puede conducir a un reporte deficiente en la literatura sobre TM¹⁰⁰.

El impacto que han tenido las guías de reporte en la mejora de los informes de los ECAs a lo largo del tiempo ha sido evaluado en diferentes ámbitos clínicos. Los resultados de dichas evaluaciones son heterogéneos. En un análisis detallado sobre el impacto de la guía CONSORT, parece que los estudios realizados sobre ECAs que evaluaron intervenciones farmacológicas^{47,111-113} reportan un mayor impacto que aquellos sobre ECAs que evaluaron intervenciones no-farmacológicas¹¹⁴⁻¹¹⁶. Sin embargo, tanto el uso de estas guías por parte de los autores como la evaluación de estudios a través de ellas no es óptimo. Recientemente se ha evidenciado que estas guías se utilizan frecuentemente de forma inadecuada¹¹⁷. Parte del problema reside en que las instrucciones sobre cómo deben usarse por parte de los autores son inconsistentes en las diferentes revistas¹¹⁸. Por otro lado, aunque actualmente existen guías y extensiones para dar cobertura a cualquier tipo de estudio, la sobrecarga que supone a los autores la estricta adhesión a ellas es alta¹¹⁹.

Los miembros del grupo CONSORT ampliaron y modificaron la declaración CONSORT original para su aplicación en diversas áreas clínicas y/o en aquellos ECAs con particulari-

dades metodológicas específicas. El resultado de este trabajo se conoce como las extensiones CONSORT (<http://www.consort-statement.org/extensions>). En 2008 se publicó la extensión CONSORT para el reporte de ECAs en intervenciones no-farmacológicas (CONSORTnpt). Los tratamientos no-farmacológicos a los que se dirige esta extensión incluyen cirugía, procedimientos técnicos, dispositivos, rehabilitación, psicoterapia, intervenciones conductuales y medicina complementaria y alternativa¹²⁰. Como el resto de extensiones, esta guía añadió a la original aspectos específicos necesarios para el reporte de los ECAs en estos campos. Aunque en el pasado se observó que la calidad del reporte no puede tomarse como un marcador directo de la calidad del ECA¹²¹, en el campo de la rehabilitación y la TM, la asociación entre pobre calidad y mal reporte ha sido determinada en varias publicaciones^{40,53,100,102}. Sin embargo, en ninguna de estas publicaciones se ha evaluado de forma directa la calidad del reporte de los ECAs en TM usando la extensión CONSORTnpt cuando, *a priori*, es la extensión más adecuada para tal propósito.

La extensión CONSORTnpt es la guía de elección para la evaluación del reporte de las TMs. La extensión original¹²⁰, fue posteriormente actualizada en 2017¹²². Uno de los elementos más destacados de dicha actualización fue la incorporación de varios ítems extraídos de otra guía de reporte, la guía TIDieR (*Template for Intervention Description and Replication*)¹²³. La guía TIDieR se publicó en 2014 con la intención de mejorar de forma específica el reporte de las intervenciones. El objetivo principal fue mejorar la descripción de las intervenciones en ECAs y por ello la extensión CONSORTnpt incorporó algunos de sus ítems en la última actualización. Sin embargo, TIDieR puede considerarse una herramienta útil para describir cualquier tipo de intervención no farmacológica independientemente del diseño metodológico utilizado. Para las ICs y, muy especialmente para la TM, la guía TIDieR constituye una herramienta muy valiosa para mejorar la calidad del reporte de las intervenciones¹²⁴, uno de sus puntos más discutidos y criticados.

Otro de los aspectos que merece especial atención en relación al reporte de los ECAs, corresponde a la comunicación de los resultados obtenidos. Los resultados proporcionan información clave sobre los efectos de las intervenciones. Por lo tanto, una correcta y detallada descripción es de suma importancia. No obstante, información clave sobre cómo se seleccionaron, definieron, midieron y analizaron los resultados de los ECAs a menudo falta o se reporta de manera deficiente^{125,126}. Sin un reporte claro y completo de los resultados, los investigadores no pueden evaluar, consolidar o replicar los hallazgos de manera adecuada, lo que dificulta la translación de la evidencia a la práctica clínica. Aunque CONSORT y SPIRIT (<https://www.spirit-statement.org>) ofrecen una guía general sobre cómo reportar los resultados, esta descripción precisa de más detalle¹²⁵⁻¹²⁸. Sin embargo, actualmente no existe una guía específica que los autores puedan seguir para describir adecuadamente los resultados de los ensayos, tanto en los protocolos como en los informes finales publicados. Para abordar este problema, un grupo internacional de expertos está desarrollando la herramienta *Instrument for reporting Planned Endpoints in Clinical Trials* (InsPECT) (<https://www.inspect-statement.org>). El grupo de investigadores que desarrollan InsPECT, liderado por la Dra. Nancy Butcher, tienen como objetivo estandarizar y armonizar el reporte de resultados de los ECAs y ya se han realizado los primeros pasos a tal efecto^{125,126}.

1.3 Justificación de la Tesis

En los apartados anteriores de esta introducción se ha descrito el escenario clínico e investigador en el que se desarrollan y evalúan las TMs. A modo de sumario, podemos decir que:

- I. Las TMs son ICs no estandarizadas cuyo efecto terapéutico se basa en la interacción de varios componentes técnicos aplicados con un alto grado de personalización hacia el paciente y su curso clínico.
- II. Pese a que actualmente los mecanismos de acción de las TMs están bien descritos desde el punto de vista biológico, el efecto clínico de dichas intervenciones está fuertemente influenciado por efectos no específicos de tipo contextual y perceptivo.
- III. La investigación relacionada con las TMs ha aumentado considerablemente en los últimos 20 años. Sin embargo, tanto la calidad metodológica como del reporte de los ECA en TM, aunque tiende a mejorar, es todavía baja.
- IV. Los motivos que explican esta falta de mejora en la calidad de los reportes se deben (al menos parcialmente) a las dificultades metodológicas existentes en la realización de ECAs considerados de alta calidad en el campo de la TM. Estas dificultades emergen de su propia naturaleza y han sido ampliamente descritas y discutidas en la literatura científica a lo largo de la historia.
- V. En la investigación en el campo de la rehabilitación y la TM se ha determinado una asociación entre la calidad del reporte y la calidad metodológica de los estudios. La guía de reporte general para la evaluación de ECAs es la guía CONSORT. Sin embargo, en el caso de intervenciones no-farmacológicas como la TM, la extensión de elección es la CONSORTnpt.
- VI. Hasta la fecha no se ha evaluado de forma exhaustiva la calidad del reporte en la literatura científica sobre TMs y el posible impacto en el tiempo que la extensión CONSORTnpt haya podido tener.

Los seis puntos anteriormente descritos plantean un escenario que, si bien ha sido ya apuntado por varios autores, no ha sido analizado en profundidad. Es por ello preciso ahondar en esta problemática y esclarecer de forma analítica cuales son las principales dificultades que se plantean. Solo a partir de la obtención de este conocimiento será posible el diseño de las estrategias apropiadas para dar respuesta a dichas dificultades y superarlas.

2. OBJETIVOS

2. Objetivos

2.1 Objetivo general

Analizar de forma exhaustiva el reporte de una amplia muestra de ECAs en TM para identificar las principales deficiencias y dificultades metodológicas que afectan negativamente a la literatura científica en este campo. La obtención de esta información persigue el fin último de desarrollar y proponer soluciones para incrementar la calidad de los estudios experimentales en TM.

2.2 Objetivos específicos

1. Describir el estado actual de la calidad del reporte de los ECAs en TM y el impacto que ha tenido la extensión CONSORTnpt desde su publicación.
2. Describir y conocer las limitaciones auto-reportadas por los investigadores en este campo y analizar las razones de dichas limitaciones.
3. Profundizar en la calidad del reporte escrito y gráfico de los resultados de los ECA en TM como elemento clave para valorar su aplicabilidad y fiabilidad.
4. Analizar la literatura científica y aportar posibles soluciones para los principales problemas metodológicos y limitaciones identificadas en nuestra revisión.

3. MÉTODOS

3. Métodos

Para abordar los objetivos específicos de la tesis, se han llevado a cabo tres estudios cuyos métodos se detallan en las respectivas publicaciones derivadas de los mismos. Dichos estudios, constituyen en su conjunto la base para presentar esta tesis a través del formato de compendio de publicaciones. A continuación, se presenta un resumen de los métodos de cada uno de los estudios.

3.1 Métodos del primer trabajo

El objetivo primario del primer trabajo consistió en evaluar la calidad del reporte de una muestra representativa de ECAs en TM publicados a texto completo en revistas indexadas. Esta evaluación del reporte se centró en aspectos específicos de la validez, la fiabilidad y el reporte de las intervenciones. A su vez, se compararon los resultados entre dos periodos de tiempo - previo y posterior a la publicación de la extensión CONSORTnpt - para evaluar el impacto de dicha guía en la calidad del reporte de los ECAs seleccionados.

3.1.1 Criterios de elegibilidad

Se incluyeron ECAs publicados en el campo de la TM. Los criterios para la inclusión fueron que al menos una de las intervenciones evaluadas en el estudio (grupo experimental o control) incluyera alguna forma de TM. Se excluyeron los artículos no escritos en inglés, aquellos con diseños distintos al ECA y los estudios que hicieran referencia a formas incompletas de diseminación como son posters y/o comunicaciones orales.

3.1.2 Estrategia de búsqueda

Se realizó una búsqueda de ECAs publicados desde el año 2000 hasta 2015 y se clasificaron en dos grupos según el año de su publicación. El grupo pre-CONSORTnpt (pre-C) estuvo compuesto por artículos publicados entre los años 2000 y 2008, y el grupo post-CONSORTnpt (post-C) por artículos publicados entre el 2009 y 2015.

La búsqueda se realizó en CENTRAL (*The Cochrane Central Register of Controlled Trials*), por ser esta la fuente más completa de ECAs^{129,130}, y la más válida para recuperar ECAs sobre fisioterapia¹³¹. La estrategia de búsqueda, basada en las recomendaciones de Pillastrini et al. (2014)¹³², se describe en el **Anexo 1**.

3.1.3 Cálculo del tamaño de la muestra

El tamaño de la muestra se calculó para determinar el número de artículos que era preciso analizar para detectar una diferencia pre-C / post-C de 4 puntos (cambio relativo de aproximadamente el 15% en el número medio de respuestas afirmativas sobre un formulario de extracción de datos original que contenía 40 elementos organizados en 25 preguntas principales). Se realizó un análisis no paramétrico mediante la prueba U de Mann-Whit-

ney. Asumimos un $\alpha=5\%$ para una aproximación bilateral y un poder estadístico mínimo del 80% (GRANMO v.7.12). Los cálculos dieron como resultado 42 artículos para cada grupo (periodo) de estudio. Decidimos aumentar a 50 artículos por grupo para mejorar la precisión ($n=100$).

3.1.4 Extracción de datos

Evaluamos el reporte de los ECAs centrándonos en los elementos más relevantes relacionados con la validez interna, la fiabilidad y la descripción de las intervenciones en este campo de estudio^{31,50,53}. Para ello, diseñamos un formulario de extracción de datos que contenía elementos seleccionados de las secciones “Métodos” y “Resultados” de la extensión CONSORTnpt, y la totalidad de la guía TIDieR. Además, completamos el formulario con elementos adicionales definidos por los investigadores. Estos elementos adicionales (no presentes en las guías originales utilizadas) estaban relacionados con las características de los ensayos y con el enmascaramiento de los participantes. Una vez diseñada la hoja de extracción de datos, tanto la comprensibilidad del formulario como el grado de acuerdo entre el grupo revisor fue evaluado mediante una prueba piloto realizada con cinco ECAs. Dicho piloto alcanzó un grado de acuerdo ‘sustancial’ entre el grupo revisor (rango 77,30% y 83,20%; valor Kappa 0,61-0,71; $P < 0,001$) según la escala de Landis¹³³.

A partir de las observaciones realizadas en la prueba piloto, el formulario de extracción de datos se optimizó en relación a su comprensibilidad y pasó a incluir finalmente un total de 44 elementos. El grupo de revisión acordó de forma más precisa los criterios de respuesta y se proporcionaron instrucciones para respaldar la tarea de los revisores (**Anexo 2**). Cada artículo fue revisado por pares, formados por el investigador principal del estudio y otro miembro del grupo de revisión. Los desacuerdos fueron discutidos y resueltos por consenso entre los pares.

3.1.5 Análisis de los datos

El análisis se realizó utilizando el paquete estadístico IBM-SPSS Statistics para Windows, versión 25.0. (Armonk, NY, Estados Unidos: IBM Corp). Las variables categóricas se presentaron descriptivamente como porcentajes. Los grupos pre-C y post-C se compararon mediante la prueba exacta de Fisher (comparación de variables dicotómicas) y la prueba de chi-cuadrado (comparación simultánea de variables categóricas).

3.2 Métodos del segundo trabajo

El objetivo principal del segundo trabajo consistió en el análisis de las limitaciones auto-reportadas (LARs) por parte de los autores en la misma muestra de ECAs en TM utilizada para el primer trabajo. Sin embargo, la muestra de estudios original formada por 100 ECAs fue ampliada con 20 ECAs adicionales publicados entre los años 2015 y el 2020, dando así una mayor cobertura temporal al análisis. Se realizó un análisis descriptivo de las LARs, se categorizaron y se discutieron las razones que los autores habían reportado para su inclusión.

3.2.1 Criterios de elegibilidad

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 Extracción de datos

Dos investigadores, uno con experiencia en metodología de investigación clínica y otro con experiencia en el campo de TM, pre-definieron por consenso una lista de 12 categorías generales de limitaciones comunes en la realización de ensayos clínicos. A partir de aquí, se desarrolló un formulario de extracción de datos que incluyó estas categorías más un campo abierto adicional para registrar cualquier otra limitación que no estuviera incluida en las categorías previamente definidas.

A los revisores se les pidió primero que identificaran (fuera en el resumen y/o en la sección de discusión de los artículos) las LARs y que las ubicaran en relación a la lista de las 12 categorías predefinidas. En caso de encontrar una LAR que no encajara en ninguna de ellas, los revisores podían registrar de igual manera este hallazgo en el campo adicional diseñado a tal efecto. A partir de aquí, la información específica contenida en los artículos fue extraída literalmente para su posterior análisis. Los datos se extrajeron por pares de revisores independientes.

3.2.3 Análisis de datos

Dos revisores compararon iterativamente las descripciones literales de las LARs para definir temas comunes. Este análisis temático acabó por definir un listado final de 24 tipos de limitaciones específicas repartidas en 12 categorías principales. Posteriormente, se realizó un análisis descriptivo con el paquete estadístico SPSS versión 22.0 (IBM Corporation, Armonk, NY). Se utilizaron frecuencias y porcentajes de variables cualitativas, incluido el intervalo de confianza establecido en el 95%

3.3 Métodos del tercer trabajo

En el tercer trabajo se realizó un análisis exhaustivo del reporte de los resultados (tanto en el texto como gráficamente) de la misma muestra de ECAs en TM utilizada para el segundo trabajo.

3.3.1 Criterios de elegibilidad

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

3.3.2 Evaluación del reporte

Reporte en el texto: Se diseñó *ad-hoc* un formulario de extracción de datos acordado por cuatro investigadores (2 expertos en metodología de la investigación y 2 expertos en TM) (**Anexo 3**). Una vez diseñado, tanto la comprensibilidad del formulario como el grado de acuerdo del grupo revisor fue evaluado mediante una prueba piloto realizada con cinco ECAs. Se observó una concordancia ‘Casi perfecta’ entre los revisores ($\kappa = 0,865$, $p < 0,05$) según la escala de Landis¹³³. En primer lugar, para cada artículo, se identificó si la variable de resultado principal del estudio se diferenciaba o no del resto de las variables reportadas. Si esta no fue declarada explícitamente, se consideró como primaria la variable utilizada para calcular el tamaño de la muestra. Si tampoco se informó la variable utilizada para el cálculo del tamaño de la muestra, se seleccionó la primera variable reportada en la sección de resultados del estudio. A partir de esta variable principal, se extrajeron datos con respecto a: a) el tipo o naturaleza de la variable de resultado; b) el seguimiento realizado (punto temporal de la medida); c) los métodos estadísticos utilizados para su cálculo; d) la medida de efecto y de precisión obtenidas (diferencias intra e inter-grupo); e) la forma en que se reportó el resultado (en el texto, gráficamente o ambos) y f) información relacionada con las pérdidas durante el seguimiento y su manejo en el análisis (imputaciones y población de análisis). También se evaluó si los artículos definían explícitamente la hipótesis del estudio (de superioridad, de no-inferioridad o de equivalencia), si incluían detalles del cálculo del tamaño de la muestra (asunciones de partida) y si aplicaban un valor mínimo de diferencia clínica importante (MCID). Todos los datos fueron extraídos por tres revisores y cualquier desacuerdo fue resuelto por un cuarto revisor. Los criterios de evaluación para cada ítem, binario o categórico, se detallan en el **Anexo 3**.

Reporte gráfico: Se identificaron aquellos ECAs que reportaron gráficamente el resultado principal. Se determinó el tipo de gráfico, la completitud y la claridad visual de acuerdo con las recomendaciones de estudios previos¹³⁴. En relación a la completitud, evaluamos si los gráficos incluían el número de participantes, títulos de gráficos y ejes, etiquetas de ejes, el significado de la varianza y si eran autoexplicativos. En relación a la claridad visual, examinamos los gráficos en busca de distorsión numérica, gráficos basura (“chart junk”), legibilidad y otros. En el **Anexo 3** se proporcionan más detalles. Todos los datos fueron extraídos por dos revisores y cualquier desacuerdo fue resuelto por un tercer revisor.

Con el fin de evaluar la evolución del reporte de los resultados a lo largo del tiempo, y debido a la ausencia de criterios validados específicamente para tal efecto en el momento de realizar dicho análisis^{125,126}, los autores acordaron un conjunto mínimo de siete categorías consideradas esenciales para un informe completo de los resultados de los ECA:

- 1) el resultado primario se declara explícitamente (y se diferencia de los resultados secundarios);
- 2) información completa sobre el cálculo del tamaño de la muestra (incluidos los valores alfa / beta y la desviación estándar o el cambio mínimo detectable);
- 3) informe completo de las pérdidas (incluidos números, motivos y criterios de imputación de los valores faltantes);
- 4) medida del efecto intragrupo (grupo experimental);
- 5) medida del efecto intragrupo (grupo de control);
- 6) medida comparativa completa entre grupos (tamaño del efecto y precisión);
- y 7) valor mínimo de diferencia clínicamente importante (MCID).

3.3.3 Análisis de datos

Se realizó estadística descriptiva utilizando frecuencias y porcentajes para variables cualitativas. Se calculó la concordancia entre cada posible par de revisores utilizando el coeficiente kappa de Cohen y el resultado se interpretó de acuerdo con la escala de Landis¹³³. Se utilizó Chi-cuadrado para comparar el informe de resultados por años y se utilizó one-way ANOVA para comparar el porcentaje de cumplimiento relacionado con las siete categorías previamente establecidas. Todos los análisis estadísticos se realizaron con SPSS versión 22.0 (IBM Corporation, Armonk, NY).

4. RESULTADOS

4. Resultados

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

1. Alvarez Gerard, Solà Ivan, Sitjà-Rabert Mercè, Fort-Vanmeerhaeghe Azahara, Gich Ignasi, Fernández Carles, Bonfill Xavier, Urrútia Gerard. ***A methodological review revealed that reporting of trials in manual therapy has not improved over time.*** Journal of Clinical Epidemiology. 20 January 2020 . doi.org/10.1016/j.jclinepi.2020.01.006. IF (2020): 6.4
2. Alvarez Gerard, Núñez-Cortés Rodrigo, Solà Ivan, Sitjà-Rabert Mercè, Fort-Vanmeerhaeghe Azahara, Fernández Carles, Bonfill Xavier, Urrútia Gerard. ***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. 32 October 2020 . doi.org/10.1016/j.jclinepi.2020.10.018. IF (2020): 6.4
3. Núñez-Cortés, Rodrigo, Alvarez, Gerard, Pérez-Bracchiglione Javier, Cabanas-Valdés Rosa, Calvo-Sanz Jordi, Bonfill Xavier, Urrútia Gerard. ***Reporting results in manual therapy clinical trials: A need for improvement.*** Int J Osteopath Med. 12 July 2021. doi:https://doi.org/10.1016/j.ijosm.2021.06.002. IF (2021): 2.14

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

La búsqueda arrojó un total de 1519 referencias de artículos pre-C y 1728 post-C que se ordenaron aleatoriamente utilizando el generador de números aleatorios de Microsoft Excel. Las primeras 50 referencias de cada grupo se seleccionaron en base a la lectura del título y el resumen, y aplicando los criterios de elegibilidad.

4.1.2 Descripción de los artículos

De los 100 artículos analizados en este primer trabajo, 77% correspondían a estudios unicéntricos (35% pre-C y 42% post-C) y 23% a estudios multicéntricos (14% pre-C y 9% post-C). El número medio de participantes fue 158 (mediana=71, SD=273, rango 6-1340) en el grupo pre-C y 78 (mediana=60, SD=60, rango 16-241) en el grupo post-C. La muestra de artículos analizados incluyó tanto artículos que evaluaron el efecto de una técnica manual aislada como aquellos que evaluaban el efecto de tratamientos manuales (que podían utilizar varias técnicas manuales y ser realizadas en distintas sesiones). Algunos estudios evaluaron paquetes terapéuticos en los que se proporcionó TM junto con otras intervenciones. En estos casos, se utilizó la información relativa a la TM utilizada para clasificar los estudios en una categoría concreta. Las intervenciones experimentales se clasificaron en seis categorías: (a) técnicas de tejidos blandos, que incluyeron estudios que evaluaron cualquiera de las siguientes modalidades: masaje, estiramiento, técnicas de energía muscular, compresión isquémica, técnicas de liberación miofascial / posicional, técnicas de "counterstrain", neuromioterapia y drenaje linfático manual; (b) movilización articular, que incluyó estudios que evaluaron cualquier técnica de movilización articular realizada en la espalda o en articulaciones periféricas (sin impulso); (c) técnicas de manipulación espinal, que incluyeron estudios que evaluaron técnicas espinales de alta velocidad y baja amplitud sin mención explícita de ninguna modalidad terapéutica específica, (d) tratamientos quiroprácticos, (e) tratamiento de manipulación osteopática y (f) acupresión o reflexología. Los estudios en las categorías (d) y (e) se asignaron respectivamente a sus categorías independientemente de si se utilizaron o no técnicas de alta velocidad y baja amplitud.

En algunos estudios se utilizó más de una intervención de control. En el grupo pre-C, diez artículos utilizaron más de un control (nueve artículos con dos controles y un artículo con tres controles) y en el grupo post-C, ocho artículos utilizaron más de un control (seis artículos con dos controles y dos artículos con tres controles).

Los estudios más frecuentes fueron los realizados con sujetos sanos / asintomáticos. En la mayoría de los casos se trataba de sujetos voluntarios reclutados en las instituciones educativas o universitarias donde se realizaron los estudios. Los problemas de salud más frecuentes fueron dolor lumbar y dolor de cuello.

El tratamiento incluyó más de una sesión en el 71% de los estudios. En el grupo pre-C, el 13% de los artículos analizaron intervenciones de sesión única y el 37% analizaron intervenciones con múltiples sesiones. En el grupo post-C, hubo 16% y 34%, respectivamente. Con respecto al seguimiento de los participantes, este se consideró a largo plazo cuando al menos una de las mediciones fue tomada 3 meses después del fin de la intervención. De acuerdo con este criterio, el 42% de los artículos reportaron un seguimiento a largo

plazo, sin diferencias significativas entre los grupos ($p = 0,31$). Otro 42% de la muestra solo reportó efectos inmediatos, 60% en el grupo post-C, comparado con 40% en el grupo pre-C.

4.1.3 Análisis comparativo del reporte

La **figura 2** (de la publicación respectiva), muestra los resultados de la comparación general entre los grupos con respecto a los elementos reportados adecuadamente. La media (SD) en el grupo pre-C fue 18,87 (4,93) frente a 19,98 (4,94) en el grupo post-C, sin diferencias estadísticamente significativas entre los dos grupos ($p = 0,26$ [3,06 a 0,85]). Con respecto a la validez y fiabilidad, solo dos ítems mostraron diferencias significativas entre los dos grupos: *"inclusión de un diagrama que muestra el flujo de participantes a través de las etapas del estudio"* ($p = 0,016$ [44.6 a 7.4]) y *"detalles de la estimación del tamaño del efecto y precisión para el resultado principal y secundarios por grupo"* ($p = 0,021$ [33 a 6.9]). Con respecto a la calidad del reporte de las intervenciones, hubo una diferencia significativa entre los dos grupos de estudio solo para el ítem *"Procedimientos"* ($p = 0,01$ [32.1 a 7.8]). Los resultados comparativos de todos los ítems que evaluaron la validez y fiabilidad así como el reporte de las intervenciones pueden encontrarse en las tablas incluidas en la publicación¹³⁵.

De forma adicional, realizamos un análisis de sensibilidad *"post-hoc"*, excluyendo los estudios publicados 2 años después de la aparición de la extensión CONSORTnpt. Este análisis adicional se hizo para evaluar el impacto de esta extensión excluyendo los artículos publicados en los años inmediatamente posteriores a su publicación. Se consideró que un análisis de este tipo permitiría analizar de forma más ecuánime la adherencia a esta extensión por parte de los autores, asumiendo que la diseminación y adopción de estas directrices requiere de algún tiempo tras su publicación inicial. Sin embargo, el análisis mostró que incluso tomando esta consideración, no hubo un cambio importante en relación a la comparación general, observándose cambios solamente en dos elementos específicos. De hecho, de los tres ítems que inicialmente mostraron diferencias estadísticas, uno dejó de serlo en este segundo análisis (*descripción significativa del tamaño del efecto y la precisión para los resultados primarios y secundarios*). Por otro lado, un ítem adicional, alcanzó dicha significación (*descripción de la experiencia del proveedor*).

4.1.4 Primera publicación

Alvarez Gerard, Solà Ivan, Sitjà-Rabert Mercè, Fort-Vanmeerhaeghe Azahara, Gich Ignasi, Fernández Carles, Bonfill Xavier, Urrútia Gerard. **A methodological review revealed that reporting of trials in manual therapy has not improved over time.** Journal of Clinical Epidemiology. 20 January 2020 . doi.org/10.1016/j.jclinepi.2020.01.006. IF (2020): 6.4

ORIGINAL ARTICLE

A methodological review revealed that reporting of trials in manual therapy has not improved over time

Gerard Alvarez^{a,b,*}, Ivan Solà^{a,c}, Mercè Sitjà-Rabert^{d,e}, Azahara Fort-Vanmeerhaeghe^{d,f}, Ignasi Gich^{c,g}, Carles Fernández^{d,e}, Xavier Bonfill^{a,c}, Gerard Urrútia^{a,c}

^aIberoamerican Cochrane Centre, Biomedical Research Institute Sant Pau, IIB Sant Pau, Barcelona, Spain

^bFoundation Centre for Osteopathic Medicine Collaboration, Spain National Centre, Barcelona, Spain

^cCIBER Epidemiología y Salud Pública (CIBERESP), Spain

^dSchool of Health Science (FCS) Blanquerna, Ramon Llull University, Barcelona, Spain

^eGlobal Research on Wellbeing (GRoW) Research Group, Ramon Llull University, Barcelona, Spain

^fFaculty of Psychology, Education Sciences and Sport (FPCEE) Blanquerna, Ramon Llull University, Barcelona, Spain

^gClinical Epidemiology and Public Health Department, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain

Accepted 15 January 2020; Published online 20 January 2020

Abstract

Objective: The aim of this review was to evaluate a selection of major reporting aspects in manual therapy (MT) trials, before and after the publication of the CONSORT extension for nonpharmacological trials (CONSORTnpt)

Study Design and Setting: We randomly selected 100 MT trials published between 2000 and 2015 and divided them into a pre-CONSORTnpt ($n = 50$) and a post-CONSORTnpt ($n = 50$) group. We extracted data on relevant issues of internal validity, reliability, and description of interventions. Two authors extracted data independently. Percentages were used for descriptive analyses, and Fisher's exact test and the chi-square test were used for group comparisons.

Results: Six different types of MT interventions with up to 20 controls were analyzed. The most common populations/conditions studied were healthy subjects and subjects with lower back or neck pain. Over 70% of studies included multi-session interventions, and 42% of studies reported long-term followup. The only significant differences between groups were the inclusion of a flowchart diagram, the estimated effect size, precision descriptions, and the description of intervention procedures.

Conclusion: Our findings suggest that trials in MT show poor reporting even after the availability of standardized guidelines. © 2020 Elsevier Inc. All rights reserved.

Keywords: CONSORT; Manual therapy; Reporting; Quality; Guidelines; Nonpharmacological

1. Background

Manual therapy (MT) is a physical treatment used by a variety of therapists to treat, mainly musculoskeletal pain and disability. It includes massage and soft tissue techniques, joint mobilization, and manipulation [1]. Since its origin, methods and approaches have evolved greatly [2].

Currently, there is evidence of the effectiveness of MT in the musculoskeletal pain field for pathologies, such as lower back pain, carpal tunnel syndrome, and knee and hip osteoarthritis, with the most robust evidence obtained when patients are classified by subgroups [3,4].

However, in spite of a growing interest in MT, there is an important lack of quality studies and high-quality evidence of its effectiveness [5], as reflected in the conclusions of numerous systematic reviews [1,6–14]. A possible explanation lies in the methodological difficulties involved in performing high-quality randomized controlled trials (RCTs) for certain medical disciplines that involve what is referred to as complex and nonpharmacological interventions [15,16]. Physical therapies in general and MT, in particular, are considered complex interventions due to the fact that they include different therapeutic modalities in the same intervention [17], assuming that the

Funding Sources: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of interest: The authors declare that they have no competing interests.

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

* Corresponding author. Iberoamerican Cochrane Center, C. Sant Antoni Maria Claret 167, Pavelló 18, Planta 0, 08025 Barcelona, Catalunya, Spain. Tel.: +34 620 214 206; fax: +34 935 537 809.

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

<https://doi.org/10.1016/j.jclinepi.2020.01.006>

0895-4356/© 2020 Elsevier Inc. All rights reserved.

What is new?**Key findings**

- Manual therapy trials show poor reporting quality.

What this adds to what was known?

- The implementation of reporting guidelines could have a minimal impact on this field.

What is the implication and what should change now?

- Guidelines specifically adapted to each kind of nonpharmacological intervention could potentially enhance reporting quality and research methods.

combination of different modalities is more effective than the sum of the parts [18–20]. Detailed reporting of such complex interventions is, therefore, crucial for the evaluation of the applicability of findings to routine practice [21].

Between 1996 and 2010, the Consolidated Standards for Reporting Trials (CONSORT) Statement and revised versions were published to significantly improve the quality of RCT reporting [22]. The CONSORT Statement has been endorsed by prominent general and specialty medical journals and leading editorial organizations. The above-said endorsement has been associated with an improvement in the quality of reporting [22–24].

In 2014, 28 rehabilitation-based journals elected to take a more aggressive stance toward the implementation of reporting guidelines in order to enhance the quality of scientific reports in this field [25]. However, recent research shows that, in spite of gradual improvement over time, there is still a need to improve both methodological quality and statistical reporting [26]. Focusing specifically on the MT literature, the evidence suggests that the quality of reporting has not improved in recent years [27], which, in turn, would suggest that the current use of CONSORT guidelines is not optimal. However, the same evidence states that these guidelines include several items with unclear definitions, leading to unreliable reporting in the MT literature [27].

In 2008, the CONSORT extension for nonpharmacological trials (CONSORTnpt), a guideline developed for psychotherapy, surgical, and rehabilitation trials, was published to respond to specific challenges that were not adequately addressed in RCT reporting [28]. However, the impact of this extension has not been assessed. The aim of this review was to evaluate a selection of major reporting aspects in manual therapy (MT) trials, before and after the publication of CONSORTnpt.

2. Methods

2.1. Eligibility criteria

RCTs published in the field of MT were included. Criteria for inclusion were that at least one of the

interventions (experimental or control) should include some form of MT. Articles not written in English, with designs other than RCTs and studies that referred to posters and oral communications, were excluded (Fig. 1).

2.2. Search strategy

We searched trials from 2000 to 2015 and classified them into two groups by year of publication. The pre-CONSORTnpt group (pre-C) was composed of articles published between 2000 and 2008, and the post-CONSORTnpt group (post-C) was composed of articles published between 2009 and 2015.

We searched the Cochrane Collaboration Trials Register (CENTRAL), as the most comprehensive source of trials [29,30], and the most valid one to retrieve trials on physical therapy [31]. The search strategy, based on the narrow search strategy proposed by Pillastrini et al. (2014) [32], is described in Appendix 1.

2.3. Sample size calculation

The sample size was calculated to determine the number of articles we needed to analyze in order to detect a 4-point pre-C/post-C difference (representing a relative change of approximately 15% in the mean number of yes responses). A nonparametric analysis was performed using the Mann-Whitney U test. We assumed an $\alpha = 5\%$ for a bilateral approximation and a minimum statistical power of 80% (GRANMO v.7.12). The calculations resulted in 42 articles for each study group, increased to 50 to improve accuracy ($n = 100$).

2.4. Data extraction

We assessed the reporting of MT RCTs focusing on the most relevant issues related to internal validity, reliability, and description of interventions [33–35]. Therefore, we designed a data extraction form to collect these issues, which contained selected items from CONSORTnpt [28], “Methods” and “Results” sections, and from the Template for Intervention Description and Replication (TIDieR) [36]. We completed the form with additional items defined by the researchers related to the trials’ characteristics and additional issues related to blinding.

We piloted the form with five RCTs reaching a substantial agreement between data extractors (range 77.30% and 83.20%; Kappa value 0.61–0.71; $P < 0.001$) [37]. After piloting, the data extraction form was optimized to include 44 items, response criteria were agreed by the review group, and instructions were provided for supporting the reviewers’ task (Appendix 2). Each article was reviewed by pairs made up of the principal investigator of the study (GA) and one other member of the review group (AF, MS, CF, GU, IS). Disagreements were discussed and resolved by the pairs.

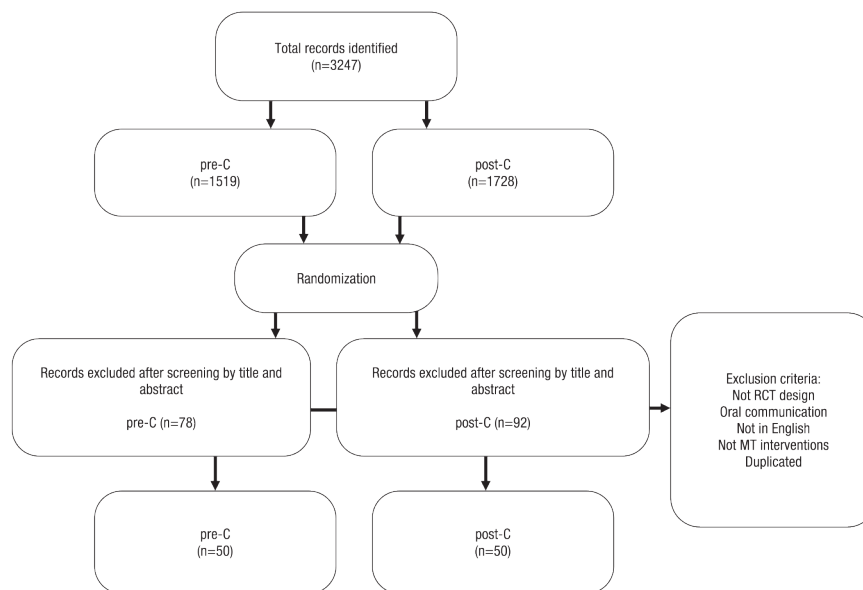


Fig. 1. Study selection flowchart.

2.5. Data analysis

The analysis was performed using IBM-SPSS Statistics for Windows, Version 25.0. (Armonk, NY, USA: IBM Corp). Categorical variables were descriptively presented as percentages. The pre-C and post-C groups were compared using the Fisher exact test (comparison of dichotomous variables) and the chi-square test (simultaneous comparison of categorical variables).

3. Results

3.1. Search results

The search yielded 1,519 pre-C and 1,728 post-C article references that were randomly ordered using the Microsoft Excel random number generator. The first 50 references in each group were selected based on a reading of the title and abstract, and by applying the eligibility criteria (Fig. 1).

3.2. Article description

Of the 100 articles analyzed, 77 corresponded to single-center studies (35 pre-C and 42 post-C) and 23 to multi-center studies (14 pre-C and 9 post-C). The mean number of study participants was 158 (median = 71, SD = 273, range 6–1340) in the pre-C group and 78 (median = 60, SD = 60, range 16–241) in the post-C group. The sample of analyzed articles included both manual techniques and treatments.

Table 1 summarizes the types of manual and control interventions. Some studies evaluated therapeutic packages in which MT was provided alongside with other interventions. In these cases, the specific MT modality was used to classify the multimodal intervention studies.

Table 1. Types of experimental and control interventions ($n = 100$).

Experimental interventions	N	Pre-C	Post-C
Soft tissue techniques	31	12	19
Joint mobilization	22	11	11
Spinal manipulation techniques	21	11	10
Chiropractic treatments	13	10	3
Osteopathic manipulative treatment	6	2	4
Acupressure-reflexology	7	4	3
Total	100	50	50
Control interventions			
Usual care	23	12	11
Exercise	21	13	8
Sham intervention	18	8	10
Joint mobilization	13	2	11
No intervention	10	6	4
Medication	6	4	2
Electrotherapy	5	2	3
Spinal manipulation techniques	3	3	0
Soft tissue techniques	4	2	2
Light massage/touch	3	0	3
Sham electrotherapy	3	0	3
Chiropractic treatment	3	3	0
Advice/education	2	1	1
Placebo capsules	1	1	0
Behavioral graded activity program	1	0	1
Acupuncture	1	1	0
Functional task practice	1	1	0
Device	1	1	0
Vibration therapy	1	1	0
Classification-based cognitive functional therapy	1	0	1
Total	121	61	60

The experimental interventions were classified in six categories: (a) soft tissue techniques, which included studies that evaluated any of the following: massage, stretching, muscle energy techniques, ischemic compression, myofascial/positional release techniques, counter-strain techniques, neuromyotherapy, and manual lymphatic drainage; (b) joint mobilization, which included studies that evaluated any joint mobilization technique performed on the back or in peripheral joints (without thrust); (c) spinal manipulation therapy (SMT) techniques, which included studies that evaluated spinal high-velocity low-amplitude (HVLA) techniques without explicit mention of any specific therapeutic modality, (d) chiropractic treatments, (e) osteopathic manipulative treatment (OMT), and (f) acupressure or reflexology. The studies in (d) and (e) were allocated to these individual categories irrespective of whether or not HVLA techniques were used.

In some studies, more than one control intervention was used (Table 1). In the pre-C group, ten articles used more than one control (nine articles with two controls and one article with three controls), and in the post-C group, eight articles used more than one control (six articles with two controls and two articles with three controls).

Table 2 summarizes the details of the studied participants/conditions. The most frequent studies were those carried out with healthy/asymptomatic subjects, in most cases, volunteer subjects recruited in the educational or university

institutions, where the studies were performed. The most frequent health problems were lower back pain and neck pain.

Treatment consisted of more than one session in 71% of the studies. For the pre-C group, 13 articles analyzed single-session interventions, and 37 articles analyzed multi-session interventions. For the post-C group, there were 16 and 34, respectively. Regarding participant follow-up, the long-term followup was defined as at least one measurement 3 months after the end of the intervention. According to this criterion, 42% of the articles reported long-term followup, with no significant differences between the groups ($P = 0.31$). Another 42% of the sample only reported immediate effects, 60% in the post-C group, compared to 40% in the pre-C group.

Fig. 2 depicts the results of an overall comparison regarding adequately reported items. The mean (SD) in the pre-C group vs. the post-C group was 18.87 (4.93) vs. 19.98 (4.94), with no statistically significant overall difference between the two groups ($P = 0.26$ [−3.06 to 0.85]). Regarding internal validity and reliability, only two items showed significant differences between the two groups: “inclusion of a diagram showing participant flow through study stages” ($P = 0.016$ [−44.6 to −7.4]) and “details of the estimated effect size and precision for primary and secondary outcomes by group” ($P = 0.021$ [−33 to −6.9]). Regarding the reporting quality of interventions, there was a significant difference between the two study groups for just the item “Procedures” ($P = 0.01$ [−32.1 to −7.8]).

Table 3 shows comparative results of items evaluating internal validity and reliability. Table 4 shows comparative results of intervention descriptions.

Additionally, we performed a post-hoc sensitivity analysis, excluding studies published 2 years after the CONSORTnpt. The analysis had no major impact in relation to the overall comparison but showed changes in two specific items. From the three items that initially showed statistical differences, one was no longer a significant *description of effect size and precision for primary and secondary outcomes*, and another one reached significance *provider expertise description*. We include a detailed analysis in Appendix 3.

4. Discussion

We conducted a study to assess the changes over time in the quality of reporting in MT trials. For this purpose, we obtained data on how trials described relevant issues of validity, reliability, and the intervention description according to items mainly captured from reporting guidelines and compared two time periods to assess the impact of one of them, that is, CONSORTnpt. Overall, our results indicate that the methodological quality of MT trial reporting is poor and has not improved over time, except for a few

Table 2. Types of participants/conditions ($n = 100$)

Participants/conditions	N	Pre-C	Post-C
Healthy/asymptomatic	12	4	8
Lower back pain	8	5	3
Acute	5	5	0
Chronic	9	4	5
Neck pain	6	4	2
Acute (including acute whiplash)	5	2	3
Chronic	4	2	2
Pregnancy/gynecological problems	7	3	4
Children- and infant-related disorders	6	2	4
Shoulder-related disorders	4	2	2
Osteoarthritis	4	1	3
Cardiac and vascular problems	3	1	2
Stroke-related disorders	3	2	1
Headache	3	1	2
Postoperative conditions	3	2	1
Spinal pain	2	1	1
Sports-related conditions	2	2	0
Craniofacial disorders	2	1	1
Plantar heel pain	2	1	1
Lateral epicondylitis	2	0	2
Other	8	5	3
Total	100	50	50

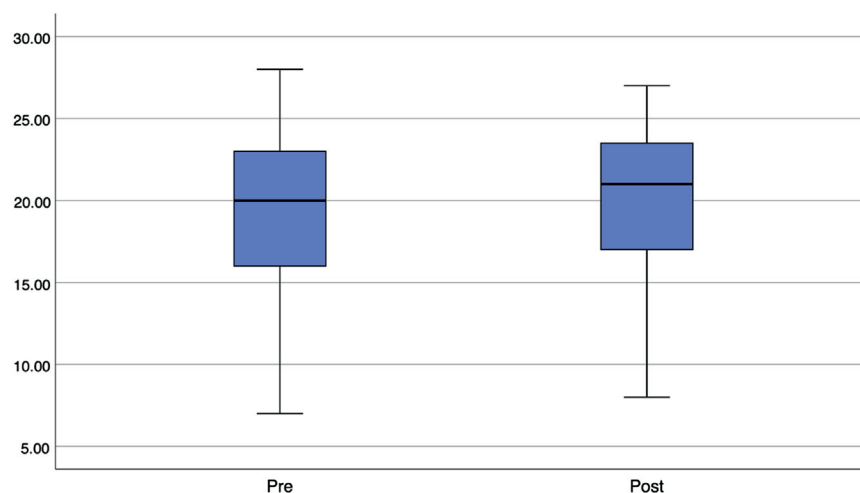


Fig. 2. Overall comparison for items adequately reported by the pre-C and post-C groups.

items. These findings could suggest that reporting guidelines have not had a significant impact on reporting quality years after its publication. Some issues have been highlighted as relevant to improve both the quality and reporting of MT research [33–35].

Our results show that despite some improvements in participant, practitioner, and setting descriptions in trials of the post-C group, there is still a need to provide more detailed descriptions. In MT studies, the description of the interventions is especially important, and guidelines, such as TIDieR [36], have been designed to improve this reporting [21]. In fact, CONSORTnpt has included most of its items in its update [38]. Intervention descriptions need to be completed very carefully in order to allow intervention evaluation and replication, as otherwise, the relevance of improving patient care is reduced, and resources are wasted [39,40]. Our results show an improvement in the description of procedures (Table 4).

In almost 60% of the analyzed MT trials, information on sample size calculation was not available, with no differences between the pre-C and post-C groups. MT trials are often based on small samples, which, in turn, usually results in type II error caused by low statistical power [33]; note that only 33% of our sample had $n \geq 100$ (38% pre-C and 28% post-C). The mean number of participants was 118, and the comparison between pre-C and post-C groups points to a falling trend in the median sample size. If we look at the literature on interventions for chronic lower back pain, the number of RCTs has risen exponentially over the past 30 years, and still there does not appear to be a corresponding increase in sample size [41]. For SMT used to treat lower back and neck pain, Rubinstein et al. [42] found that although the sample size appeared to increase over time, it did not do so to a statistically significant degree (overall $P = 0.79$). The same authors also found a linear trend over time in the odds of sample size being calculated a priori, a fact that may point to an improvement in this particular aspect.

Since nonpharmacological interventions, such as MT, are typically used to treat chronic conditions [43–45], treatment may last several weeks, so outcome evaluation of necessity includes long-term followup. Trialists, therefore, need to be aware of the fact that a small or underpowered sample in an RCT with long-term followup can potentially be affected by attrition bias. Information on attrition and its impact on baseline imbalances have been proposed as an item for inclusion in CONSORT guidelines [46], but no specific mention has been included in the updated version of the CONSORTnpt [38].

Regarding randomization and allocation concealment, full information was provided in 64% of the articles, with no differences between groups. We argue that this is an easy improvement to apply, as it poses none of the methodological challenges posed by other trial design requirements.

Blinding is critical to all nonpharmacological interventions [47] and is considered to be a major challenge in rehabilitation and physical medicine trials [33–35,48,49]. A modification included in the updated 2017 version of the CONSORTnpt (point 11c) recognizes this major difficulty and recommends to at least describe attempts to limit bias when blinding is not possible [38]. Table 5 confirms that blinding in MT trials is an aspect that has hardly improved over the years, with a favorable trend only evident for outcome evaluator blinding. Improvements in blinding strategies are probably linked to the use of better controls (sham or placebo interventions), another aspect, which MT trials need to improve [50,51]. Moreover, knowing the crucial role of contextual factors in nonpharmacological intervention effects [52–54] and bearing in mind that blinding success is rarely tested [55], attempts to assess whether masking was effective should be proposed as a good methodological practice for MT trials [56].

In our sample, 59% of the articles included a flowchart diagram and a statistically significant difference was found

Table 3. Comparative results related to internal validity and reliability reporting items

Description	% Total YES	NA	Pre-C YES	Post-C YES	% Change	P	95% CI
Eligibility criteria for participants are stated	97%	—	47 (94%)	50 (100%)	6%	0.61	[−12 to 0.5]
Eligibility criteria for settings and locations are stated (single-center)	51%	—	22 (44%)	29 (58%)	14%	0.31	[−33.4 to 5.4]
Eligibility criteria for settings and locations are stated (multicenter)	15%	77%	9 (18%)	6 (12%)	−6%	0.32	[−7.9 to 19.9]
Eligibility criteria for persons performing the intervention are stated	56%	—	23 (46%)	33 (66%)	20%	0.07	[−39.1 to 0.9]
Objectives and hypotheses are stated	100%	—	50 (50%)	50 (50%)	0%	—	—
Primary outcome measures are clearly defined	47%	—	24 (48%)	23 (46%)	−2%	0.84	[−17.6 to 21.6]
Methods are included that enhanced measurement quality	82%	4%	40 (80%)	42 (84%)	4%	0.51	[−19 to 11]
How sample size was determined is indicated and, when applicable, details are provided of whether and how clustering by care provider or center was done	41%	—	17 (34%)	24 (48%)	14%	0.22	[−33.1 to 5]
A random allocation sequence method was used	64%	—	29 (58%)	35 (70%)	12%	0.40	[−30.7 to 6.6]
Care provider allocation to each trial group is described	4%	47%	2 (4%)	2 (4%)	0%	—	[−7.6 to 7.6]
The method used to implement the random allocation sequence is described	51%	—	22 (44%)	29 (58%)	14%	0.31	[−33.4 to 5.4]
The persons who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions are indicated	36%	—	16 (32%)	20 (40%)	8%	0.53	[−26.8 to 10.8]
Participant blinding to group allocation is indicated	32%	—	14 (28%)	18 (36%)	8%	0.52	[−26.2 to 10.2]
Whether the persons performing the intervention were blinded to group allocation is indicated	3%	—	2 (4%)	1 (2%)	−2%	0.61	[−4.6 to 8.6]
Whether the persons evaluating outcomes were blinded to group allocation is indicated	60%	—	26 (52%)	34 (68%)	16%	0.22	[−34.9 to 2.9]
Whether the persons performing co-interventions were blinded to group allocation is indicated	10%	68%	5 (10%)	5 (10%)	0%	0.82	[−11.8 to 11.8]
Statistical methods to compare groups for primary outcomes are described	100%	—	50 (50%)	50 (50%)	0%	—	—
Methods for additional analyses (subgroup analyses and adjusted analyses) are described	39%	41%	21 (42%)	18 (36%)	−6%	0.73	[−13.1 to 25.1]
Details of whether and how clustering by care provider or center was done (when applicable) are provided	3%	80%	2 (4%)	1 (2%)	−2%	0.10	[−4.6 to 8.6]
Participant flow through stages is described	59%	—	23 (46%)	36 (72%)	26%	0.016 ^a	[−44.6 to −7.4]
The number of randomly allocated participants is indicated	100%	—	50 (50%)	50 (50%)	0%	—	—
The number of participants receiving the intended treatment is indicated	97%	—	48 (96%)	49 (98%)	2%	1.0	[−8.6 to 4.6]

(Continued)

Table 3. Continued

Description	% Total YES	NA	Pre-C YES	Post-C YES	% Change	P	95% CI
The number of participants completing the study protocol is indicated	85%	—	42 (84%)	43 (86%)	2%	1.0	[−16 to 12]
The number of participants analyzed for the primary outcome is indicated	95%	—	48 (96%)	47 (94%)	−2%	0.36	[−6.5 to 10.5]
Protocol deviations from study as planned are indicated	5%	30%	2 (4%)	3 (6%)	2%	0.58	[−10.5 to 6.5]
The number of care providers and centers performing the intervention in each group is indicated	34%	—	15 (30%)	19 (38%)	8%	0.53	[−26.5 to 10.5]
The number of patients treated by each care provider and center is indicated	17%	—	8 (16%)	9 (18%)	2%	0.83	[−16.7 to 12.7]
Losses and exclusions after randomization are indicated	89%	—	44 (88%)	45 (90%)	2%	1.0	[−14.3 to 10.3]
Reasons for losses and exclusions after randomization are described	51%	—	25 (50%)	26 (52%)	2%	0.99	[−21.6 to 17.6]
The dates defining recruitment are indicated	52%	—	22 (44%)	30 (60%)	16%	0.16	[−35.3 to 3.3]
The dates defining follow-up are indicated	71%	—	37 (74%)	34 (68%)	−6%	0.66	[−11.7 to 23.7]
The baseline demographic and clinical characteristics of each group are provided	91%	—	43 (86%)	48 (96%)	10%	0.31	[−21 to 10.5]
The number of participants (denominator) for each group included in each analysis is indicated	97%	—	48 (96%)	49 (98%)	2%	1.0	[−8.6 to 4.6]
Whether or not analysis was by intention-to-treat is indicated	34%	—	18 (36%)	16 (32%)	−4%	0.67	[−14.6 to 22.6]
For each primary and secondary outcome, a summary of results for each group and the estimated effect size and its precision are described	86%	—	38 (76%)	48 (96%)	20%	0.021 ^a	[−33 to −6.9]
Other analyses are described, including subgroup analyses and adjusted analyses, with an indication that they are prespecified or exploratory	38%	53%	22 (44%)	16 (32%)	−12%	0.37	[−6.8 to 30.9]
All important harms and unintended effects in each group are indicated	47%	—	22 (44%)	25 (50%)	6%	0.69	[−25.5 to 13.5]

NA, not applicable.

^a $P < 0.05$.

between the two study groups. As shown in Table 6, the number of randomly allocated participants and information on missing data, dropouts, and losses to followup were generally well reported.

Another important item is “stating the number of participants included in each analysis and whether the analysis was by intention-to-treat” (ITT). Although the former was reported in 97% of the articles, only 34% of articles referred to ITT analyses, which, in fact, is information that is infrequently included in publications on physical therapies [57]. Trials can fail to meet this criterion by not using ITT analysis, by not reporting that ITT analysis was used or by misinterpreting the definition of ITT analysis. Furthermore, full follow-up and equal final group sizes do not guarantee the quality of ITT analyses [58]. Given intrinsic MT characteristics, RCTs often aim to evaluate the

effectiveness (rather than the efficacy) of interventions, typically by applying a pragmatic approach to the design. It has been suggested that the result analysis should apply the same pragmatic approach, that is, reflecting clinical practice as faithfully as possible, as the ITT analysis should provide an indication of the clinical benefits of the treatment [59]. For the above reasons, we suggest that inclusion in the published report of a formal statement regarding the type of analysis should be strongly recommended. While data can be imputed alongside an ITT analysis, the fact remains that none of any original articles on ITT analysis, the Cochrane Handbook, the CONSORT statement, or the PEDro Scale specifically recommend, including this information [58]. In our sample, the imputation of missing data was described in just half of the 66% of articles that described dropout rates.

Table 4. Comparative results related to intervention description reporting items

Description	% Total YES	NA	Pre-C YES	Post-C YES	% Change	P	95% CI
Provide the name or a phrase that describes the intervention	99%	—	48 (96%)	50 (100%)	4%	0.49	[−9.4 to 1.4]
Describe any rationale, theory, or goal of the elements essential to the intervention	99%	—	48 (96%)	50 (100%)	4%	0.49	[−9.4 to 1.4]
Materials: Describe any physical or informational materials used in the intervention, including those provided to participants or used in intervention delivery or in training of intervention providers. Provide information on where the materials can be accessed (such as online appendix, URL)	100%	—	—	—	—	—	—
Procedures: Describe each of the procedures, activities, and/or processes used in the intervention, including any enabling or support activities	88%	—	39 (78%)	49 (98%)	20%	0.01 ^a	[−32.1 to −7.8]
For each category of intervention provider (such as psychologist, nursing assistant), describe their expertise, background, and any specific training given	53%	—	22 (44%)	31 (62%)	18%	0.16	[−37.2 to 1.2]
Describe the modes of delivery (such as face to face or by some other mechanism, such as internet or telephone) of the intervention and whether it was provided individually or in a group	100%	—	—	—	—	—	—
Describe the type(s) of location(s) where the intervention occurred, including any necessary infrastructure or relevant features	36%	—	14 (28%)	22 (44%)	16%	0.14	[−34.6 to 2.5]
Describe the number of times the intervention was delivered and over what period of time, including the number of sessions, their schedule, and their duration, intensity, or dose	92%	—	43 (86%)	49 (98%)	12%	0.15	[−18.9 to 2.2]
If the intervention was planned to be personalized, titrated or adapted, then describe what, why, when, and how	34%	30%	17 (34%)	17 (34%)	0%	0.47	[−18.6 to 18.6]
If the intervention was modified during the course of the study, describe the changes (what, why, when, and how)	5%	89%	1 (2%)	4 (8%)	6%	0.27	[−14.5 to 2.4]
Planned: If intervention adherence or fidelity was assessed, describe how and by whom, and if any strategies were used to maintain or improve fidelity, describe them	23%	66%	11 (22%)	12 (24%)	2%	0.58	[−18.5 to 14.5]
Actual: If intervention adherence or fidelity was assessed, describe the extent to which the intervention was delivered as planned	20%	70%	10 (5%)	10 (5%)	0%	0.74	[−15.7 to 15.7]

NA, not applicable.

^a $P < 0.05$.

“Information on effect size and precision by group” was provided in 86% of our sample, with a significant improvement in post-C articles. Nonetheless, we would

suggest that intergroup comparison rather than intragroup comparison should be specified in the results reported in MT trials.

Table 5. Reporting of blinding ($n = 100$)

Who was blinded	Total YES	NA	Pre-C YES	Post-C YES	% Change	P	95% CI
Participants	31%	—	13 (26%)	18 (36%)	10%	0.52	[−28 to 8]
Practitioners	3%	—	2 (4%)	1 (2%)	−2%	0.61	[−4.6 to 8.6]
Outcome evaluators	60%	—	26 (52%)	34 (68%)	16%	0.22	[−34.9 to 2.9]
Co-intervention practitioners	10%	68%	5 (10%)	5 (10%)	—	0.82	[−15.7 to 15.7]

NA, not applicable.

Finally, adverse events were mentioned in 47% of all trials, with no differences between groups. This low rate of reported adverse events is possibly due to no adverse events occurring in the MT trials analyzed. While the scientific literature is consistent in describing MT as a safe intervention [60–62], adverse events have been reported, ranging from catastrophic (cervical artery dissection resulting in a stroke) for SMT to mild, for example, transient muscle soreness or stiffness, considered an expected outcome of treatment [60,63]. The reporting of adverse events should be considered mandatory regardless of their gravity or frequency. However, and although improvement is evident since the CONSORT guidelines became available [64],

for SMT, in particular, the current level of reporting of adverse events can only be deemed inadequate [65].

For our study of randomly selected articles from CONSORT-adherent and non-CONSORT-adherent journals, our findings are in line with those of Riley et al. [27] for only CONSORT-adherent journals. A similar study of chiropractic treatment that explored associations between certain variables and overall reporting quality found that reporting was influenced by year of publication and sample size, but not by journal type, funding source, or positive outcomes [66].

While our results are also consistent with other similar reviews of physical therapies [26,67], they contrast with

Table 6. Reporting of participant flow and follow-up details ($n = 100$)

Total YES	NA	Pre-C yes	Post-C yes	% Change	P	95% CI
Is the number of participants who were randomly assigned indicated?						
100%	—	50 (100%)	50 (100%)	—	—	—
Is the number of participants who received the intended treatment indicated?						
97%	—	48 (96%)	49 (98%)	2%	1.0	[−8.6 to 4.6]
Is compliance with the intended treatment described?						
85%	—	42 (84%)	43 (86%)	2%	1.0	[−16 to 12]
Are missing data/dropouts/losses to follow-up described?						
89%	—	44 (88%)	45 (90%)	2%	1.0	[−14.3 to 10.3]
Are reasons given for missing data/dropouts/losses to follow-up?						
51%	37%	25 (50%)	26 (52%)	2%	0.99	[−21.6 to 17.6]
Is the number of participants analyzed for the primary outcome indicated?						
95%	—	48 (96%)	47 (94%)	−2%	0.36	[−6.5 to 10.5]
Have there been any deviations from the planned protocol?						
5%	30%	2 (4%)	3 (6%)	2%	0.58	[−10.5 to 6.5]
Is the number of therapists/centers given for each intervention?						
34%	—	15 (30%)	19 (38%)	8%	0.53	[−26.5 to 10.5]
Is the number of patients treated by each therapist/center given?						
17%	33%	8 (16%)	9 (18%)	2%	0.83	[−16.7 to 12.7]

NA, not applicable.

positive results of reporting quality found in other fields [22,24,68,69]. This would suggest that the application of the CONSORT guidelines may have a different impact depending on the study field. Although an association between poor methodological quality in design and poor reporting quality is to be expected, results to date and the comments of a number of authors raise questions regarding the use of reporting guidelines and their effectiveness in improving the scientific literature:

4.1. Are the items in the CONSORT guidelines sufficiently clear for manuscript authors?

The availability of resources to enhance understanding of and accessibility to reporting guidelines and their items has been discussed [70,71]. Moreover, specific instructions on how CONSORT should be used by authors are inconsistent across journals and publishers [72]. To the best of our knowledge, there are no specific study surveying authors regarding the clarity of CONSORT checklist items. On the other hand, studies evaluating strategies aiming to help authors prepare more complete trial reports (e.g., WebCONSORT) conclude that more detailed information on how to implement each item within the context of a specific trial is needed [73]. As an example, our study involved a review team comprised of clinical MT experts and researchers with extensive experience in methodological aspects of RCTs. While the pilot run of the reviewers' analysis resulted in substantial agreement, clear differences in criteria arose and needed to be discussed. The use of a structured approach for reporting research [74] or the use of the writing aid tool COBWEB [75] are considered good strategies to help authors improve adherence to reporting guidelines in health research [70].

4.2. For CONSORT-adherent journals, is there a standardized process to evaluate article adherence to the reporting guidelines?

Beyond general recommendations, a mechanism needs to be put in place so that article submission is conditional on the inclusion of the information required by the guidelines [72,76]. Moreover, researchers and peer-reviewers should also receive training on the use of these guidelines [76,77], and editors should incorporate them into the editorial process, as inconsistencies between what authors claimed on submitted checklists and what was actually reported in the published paper have been found [76]. This requirement would not only enhance reporting quality but could also potentially improve the methodological quality of studies.

4.3. Is it possible that the usefulness of CONSORT guidelines depends on the study field? Is the extension for nonpharmacological interventions suitable for all nonpharmacological interventions?

To date, the number of studies pointing to a significant impact of the CONSORT guidelines (or its extensions) on

pharmacological interventions [22,69,78] seems to be greater than that in nonpharmacological interventions [79–81]. The nonpharmacological label may be an umbrella term that covers a wide number and variety of very different interventions. It is plausible to consider that the same set of guidelines may not adequately cover the specific features of different kinds of interventions. For instance, the CONSORT guidelines have had a limited impact on reporting quality in the fields of dentistry and surgery [79–83], while the Standards for Reporting Interventions in Clinical Trials of Acupuncture (STRICTA) [84] have had a positive impact on the quality of clinical trial reporting in acupuncture [85–87], despite some poor results [88–90]. It could be argued that the existence of guidelines or extensions adapted to specific interventions could potentially enhance reporting quality and research methods. However, a new strategy for developing guidelines or extensions should be considered in order to reduce authors' potential workload and to increase adherence [71].

4.4. Strengths and limitations

We acknowledge several limitations in this review. The selected articles may not be representative of the complete MT body of evidence. However, we have tried to minimize possible selection bias by using a random sample of published trials. Furthermore, we searched CENTRAL as the most comprehensive source of trials [29,30], and the most valid one to retrieve trials on physical therapy [31]. Although we initially searched MEDLINE, we realized that this database missed relevant journals in the MT field that are included in CENTRAL.

On the other hand, we did not differentiate between CONSORT-adherent and non-CONSORT-adherent journals in selecting the sample, so this could have an impact on our findings. Also, our sample includes MT RCTs published up to 2015, so our study is not able to assess if the quality of reporting has changed in the last 4 years. For example, taking into account that TIDieR was published in 2014 and that it has been recommended for its use in physical [91] and manual therapies [21], it is reasonable to expect an effect on most recent publications related to intervention descriptions.

We assessed the evolution of reporting in MT trials from a comprehensive approach focusing on relevant issues in MT regarding internal validity, reliability, and the description for the interventions, which are identified as shortcomings in this field. We designed a data extraction form selecting items from standardized reporting guidelines (i.e., CONSORTnpt and TIDieR), and we also included others that were considered relevant by the researchers conducting the study. Some of these aspects have been added to the CONSORTnpt update and indirectly validates our decision.

In relation to the reviewer response options, while strategies were established to minimize differences in criteria, results may have varied among our review group members.

Our reviewers reported some difficulties in determining how to rate some items, in some cases due to a lack of/deficient information in the articles themselves and, in other cases, due to the lack of response options—binary in some cases (YES/NO) and in other cases, including NOT APPLICABLE. The fact that a NOT CLEAR option was not included for the YES/NO statements left the reviewers having to opt for either response in cases of doubt.

5. Conclusion

Reporting guidelines aim to help researchers better describe their research to ultimately improve methodological transparency and thus allowing replication and bias assessment. Regarding RCTs, authors are required to present their research in compliance with CONSORT guidelines and most peer-reviewed journals that endorse CONSORT. However, several years after the publication of CONSORT guidelines and extensions, the quality of reporting continues to be very uneven. Our findings suggest that trials in MT show poor reporting even after the availability of standardized guidelines. Adherence to these reporting guidelines should be a mandatory submission requirement for MT trialists. In relation to nonpharmacological interventions, the development of extensions related to specific interventions could potentially improve the quality of reporting.

CRedit authorship contribution statement

Gerard Alvarez: Conceptualization, Methodology, Investigation, Writing - original draft, Writing - review & editing. **Ivan Solà:** Supervision, Investigation, Writing - review & editing. **Mercè Sitjà-Rabert:** Investigation, Writing - review & editing. **Azahara Fort-Vanmeerhaeghe:** Investigation, Writing - review & editing. **Ignasi Gich:** Formal analysis, Writing - review & editing. **Carles Fernández:** Investigation, Writing - review & editing. **Xavier Bonfill:** Supervision, Writing - review & editing. **Gerard Urrútia:** Conceptualization, Methodology, Supervision, Writing - review & editing.

Acknowledgments

The authors wish to acknowledge Ms. Victoria Leo for her contribution in improving the manuscript.

Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.jclinepi.2020.01.006>.

References

- [1] French HP, Brennan A, White B, Cusack T. Manual therapy for osteoarthritis of the hip or knee - a systematic review. *Man Ther* 2011; 16(2):109–17.
- [2] Rivett DA. Editorial - manual therapy cults. *Man Ther* 1999;4(3): 125–6.
- [3] Bialosky JE, Bishop MD, Price DD, Robinson ME, George SZ. The mechanisms of manual therapy in the treatment of musculoskeletal pain: a comprehensive model. *Man Ther* 2009;14:531–8.
- [4] Voogt L, de Vries J, Meeus M, Struyf F, Meuffels D, Nijs J. Analgesic effects of manual therapy in patients with musculoskeletal pain: a systematic review. *Man Ther* 2015;20(2):250–6.
- [5] Clar C, Tsertsvadze A, Court R, Hundt G, 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(1):12.
- [6] Hondras MA, Linde K, Jones AP. Manual therapy for asthma. *Cochrane Database Syst Rev* 2002;2:CD001002.
- [7] Bokarius AV, Bokarius V. Evidence-based review of manual therapy efficacy in treatment of chronic musculoskeletal pain. *Pain Pract* 2010;10(5):451–8.
- [8] Orrock PJ, Myers SP. Osteopathic intervention in chronic non-specific low back pain: a systematic review. *BMC Musculoskelet Disord* 2013;14:129.
- [9] Page MJ, Green S, Kramer S, Johnston RV, Mcbain B, Chau M, et al. Manual therapy and exercise for adhesive capsulitis. *Cochrane Database Syst Rev* 2014;8:1–225.
- [10] Peek AL, Miller C, Heneghan NR. Thoracic manual therapy in the management of non-specific shoulder pain: a systematic review. *J Man Manip Ther* 2015;23(4):176–87.
- [11] Cicchitti L, Martelli M, Cerritelli F. Chronic inflammatory disease and osteopathy: a systematic review. *PLoS One* 2015;10:e0121327.
- [12] Ruffini N, D'Alessandro G, Cardinali L, Frondaroli F, Cerritelli F. Osteopathic manipulative treatment in gynecology and obstetrics: a systematic review. *Complement Ther Med* 2016;26:72–8.
- [13] Cerritelli F, Ruffini N, Lacorte E, Vanacore N. Osteopathic manipulative treatment in neurological diseases: systematic review of the literature. *J Neurol Sci* 2016;369:333–41.
- [14] Cerritelli F, Lacorte E, Ruffini N, Vanacore N. Osteopathy for primary headache patients: a systematic review. *J Pain Res* 2017;10:601–11.
- [15] 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.
- [16] Clark AM. What are the components of complex interventions in healthcare? Theorizing approaches to parts, powers and the whole intervention. *Soc Sci Med* 2013;93:185–93.
- [17] 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.
- [18] Hawe P, Shiell A, Riley T. Complex interventions: how “out of control” can a randomised controlled trial be? *Br Med J* 2004;328(7455): 1561–3.
- [19] Dieppe P. Complex interventions. *Musculoskeletal Care* 2004;2(3): 180–6.
- [20] Coronado RA, Bialosky JE. Manual physical therapy for chronic pain: the complex whole is greater than the sum of its parts. *J Man Manip Ther* 2017;25(3):115–7.
- [21] 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.
- [22] Plint AC, Moher D, Morrison A, Schulz K, Altman DG, Hill C, et al. Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. *Med J Aust* 2006;185:263–7.
- [23] Reveiz L, Chapman E, Asial S, Munoz S, Bonfill X, Alonso-Coello P. Risk of bias of randomized trials over time. *J Clin Epidemiol* 2015; 68:1036–45.
- [24] Dechartres A, Charles P, Hopewell S, Ravaud P, Altman DG. Reviews assessing the quality or the reporting of randomized controlled

- trials are increasing over time but raised questions about how quality is assessed. *J Clin Epidemiol* 2011;64:136–44.
- [25] Chan L, Heinemann AW, Roberts J. Elevating the quality of disability and rehabilitation research: mandatory use of the reporting guidelines. *Am J Occup Ther* 2014;68(2):127–9.
 - [26] Gonzalez GZ, Moseley AM, Maher CG, Nascimento DP, Luciola da Cunha Menezes C, Costa LOP. Methodologic quality and statistical reporting of physical therapy randomized controlled trials relevant to musculoskeletal conditions. *Arch Phys Med Rehabil* 2018;99:129–36.
 - [27] 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(5):241–52.
 - [28] Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. *Ann Intern Med* 2008;148:295–309.
 - [29] Egger M, Smith GD. Bias in location and selection of studies. *Br Med J* 1998;316(7124):61–6.
 - [30] Higgins J, Green S. *Cochrane Handbook for Systematic Reviews of Interventions* 2011. Available at <http://handbook.cochrane.org/>. Accessed January 5, 2020.
 - [31] Michaleff ZA, Costa LOP, Moseley AM, Maher CG, Elkins MR, Herbert RD, et al. CENTRAL, PEDro, PubMed, and EMBASE are the most comprehensive databases indexing randomized controlled trials of physical therapy interventions. *Phys Ther* 2011;91:190–7.
 - [32] Pillastri P, Vanti C, Curti S, Mattioli S, Ferrari S, Violante FS, et al. Using PubMed search strings for efficient retrieval of manual therapy research literature. *J Manipulative Physiol Ther* 2015;38:159–66.
 - [33] Koes BW. How to evaluate manual therapy: value and pitfalls of randomized clinical trials. *Man Ther* 2004;9(4):183–4.
 - [34] Wade DT, Smeets RJEM, Verbunt JA. Research in rehabilitation medicine: methodological challenges. *J Clin Epidemiol* 2010;63:699–704.
 - [35] Arienti C, Armijo-Olivo S, Minozzi S, Lazzarini SG, Patrini M, Negrini S. 60 Methodological issues in rehabilitation research: a scoping review. *BMJ Evid Based Med* 2019;24:A35.
 - [36] 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. *Gesundheitswesen* 2016;78:175–88.
 - [37] Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 2006;33:159–74.
 - [38] Boutron I, Altman DG, Moher D, Schulz KF, Ravaud P, CONSORT NPT Group, et al. CONSORT statement for randomized trials of non-pharmacologic treatments: a 2017 update and a CONSORT extension for nonpharmacologic trial abstracts. *Ann Intern Med* 2017;167:40–7.
 - [39] Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. *Lancet* 2009;374(9683):86–9.
 - [40] Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, Julious S, et al. Reducing waste from incomplete or unusable reports of biomedical research. *Lancet* 2014;383(9913):267–76.
 - [41] Henschke N, Kuijpers T, Rubinstein SM, van Middelkoop M, Ostelo R, Verhagen A, et al. Trends over time in the size and quality of randomised controlled trials of interventions for chronic low-back pain. *Eur Spine J* 2012;21(3):375–81.
 - [42] Rubinstein SM, Van Eekelen R, Oosterhuis T, De Boer MR, Ostelo RWJG, Van Tulder MW. The risk of bias and sample size of trials of spinal manipulative therapy for low back and neck pain: analysis and recommendations. *J Manipulative Physiol Ther* 2014;37(8):523–41.
 - [43] Hassett AL, Williams DA. Non-pharmacological treatment of chronic widespread musculoskeletal pain. *Best Pract Res Clin Rheumatol* 2011;25:299–309. <https://doi.org/10.1016/j.berh.2011.01.005>.
 - [44] Reimold AM, Chandran V. Nonpharmacologic therapies in spondyloarthritis. *Best Pract Res Clin Rheumatol* 2014;28:779–92.
 - [45] Chang K-L, Fillingim R, Hurley RW, Schmidt S. Chronic pain management: nonpharmacological therapies for chronic pain. *FP Essent* 2015;432:21–6.
 - [46] Hewitt CE, Kumaravel B, Dumville JC, Torgerson DJ. Assessing the impact of attrition in randomized controlled trials. *J Clin Epidemiol* 2010;63:1264–70.
 - [47] Boutron I, Guittet L, Estellat C, Moher D, Hróbjartsson A, Ravaud P. Reporting methods of blinding in randomized trials assessing non-pharmacological treatments. *PLoS Med* 2007;4(2):0370–80.
 - [48] 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.
 - [49] 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(1):6–13.
 - [50] Hancock MJ, Maher CG, Latimer J, McAuley JH. Selecting an appropriate placebo for a trial of spinal manipulative therapy. *Aust J Physiother* 2006;52(2):135–8.
 - [51] Cerritelli F, Verzella M, Cicchitti L, D'Alessandro G, Vanacore N. The paradox of sham therapy and placebo effect in osteopathy: a systematic review. *Medicine (Baltimore)* 2016;95(35):e4728.
 - [52] Di Blasi Z, Harkness E, Ernst E, Georgiou A, Kleijnen J. Influence of context effects on health outcomes: a systematic review. *Lancet* 2001;357:757–62.
 - [53] Di Blasi Z, Kleijnen J. Context effects. Powerful therapies or methodological bias? *Eval Health Prof* 2003;26(2):166–79.
 - [54] 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(1):27.
 - [55] Hróbjartsson A, Forfang E, Haahr MT, Als-Nielsen B, Brorson S. Blinded trials taken to the test: an analysis of randomized clinical trials that report tests for the success of blinding. *Int J Epidemiol* 2007;36:654–63.
 - [56] Kolahi J, Bang H, Park J. Towards a proposal for assessment of blinding success in clinical trials: up-to-date review. *Community Dent Oral Epidemiol* 2009;37(6):477–84.
 - [57] Moseley AM, Herbert RD, Maher CG, Sherrington C, Elkins MR. Reported quality of randomized controlled trials of physiotherapy interventions has improved over time. *J Clin Epidemiol* 2011;64:594–601.
 - [58] Elkins MR, Moseley AM. Intention-to-treat analysis. *J Physiother* 2015;61(3):165–7.
 - [59] Sedgwick P. Intention to treat analysis versus per protocol analysis of trial data. *BMJ* 2015;350:h681.
 - [60] Carnes D, Mars TS, Mullinger B, Froud R, Underwood M. Adverse events and manual therapy: a systematic review. *Man Ther* 2010;15:355–63.
 - [61] Paanalahti K, Holm LW, Nordin M, Asker M, Lyander J, Skillgate E. Adverse events after manual therapy among patients seeking care for neck and/or back pain: a randomized controlled trial. *BMC Musculoskelet Disord* 2014;15:77.
 - [62] Todd AJ, Carroll MT, Robinson A, Mitchell EKL. Adverse events due to chiropractic and other manual therapies for infants and children: a review of the literature. *J Manipulative Physiol Ther* 2015;38:699–712.
 - [63] Carlesso LC, Cairney J, Dolovich L, Hoogenes J. Defining adverse events in manual therapy: an exploratory qualitative analysis of the patient perspective. *Man Ther* 2011;16:440–6.
 - [64] Gorrell LM, Brown B, Lystad RP, Engel RM. Predictive factors for reporting adverse events following spinal manipulation in randomized clinical trials - secondary analysis of a systematic review. *Musculoskelet Sci Pract* 2017;30:34–41.
 - [65] Gorrell LM, Engel RM, Brown B, Lystad RP. The reporting of adverse events following spinal manipulation in randomized clinical trials—a systematic review. *Spine J* 2016;16(9):1143–51.

- [66] Karpouzis F, Bonello R, Pribicevic M, Kalamir A, Brown BT. Quality of reporting of randomised controlled trials in chiropractic using the CONSORT checklist. *Chiropr Man Therap* 2016;24:19.
- [67] 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(1):36–43.
- [68] Kane RL, Wang J, Garrard J. Reporting in randomized clinical trials improved after adoption of the CONSORT statement. *J Clin Epidemiol* 2007;60:241–9.
- [69] Turner L, Shamseer L, Altman DG, Schulz KF, Moher D. Does use of the CONSORT Statement impact the completeness of reporting of randomised controlled trials published in medical journals? A Cochrane review. *Syst Rev* 2012;1(1):60.
- [70] Blanco D, Altman D, Moher D, Boutron I, Kirkham JJ, Cobo E. Scoping review on interventions to improve adherence to reporting guidelines in health research. *BMJ Open* 2019;9(5):e026589.
- [71] Ghosn L, Boutron I, Ravaud P. Consolidated Standards of Reporting Trials (CONSORT) extensions covered most types of randomized controlled trials, but the potential workload for authors was high. *J Clin Epidemiol* 2019;113:168–75.
- [72] Shamseer L, Hopewell S, Altman DG, Moher D, Schulz KF. Update on the endorsement of CONSORT by high impact factor journals: a survey of journal “Instructions to Authors” in 2014. *Trials* 2016;17(1):301.
- [73] Hopewell S, Boutron I, Altman DG, Barbour G, Moher D, Montori V, et al. Impact of a web-based tool (WebCONSORT) to improve the reporting of randomised trials: results of a randomised controlled trial. *BMC Med* 2016;14(1):199.
- [74] Pandis N, Turpin DL. Enhancing CONSORT compliance for improved reporting of randomized controlled trials. *Am J Orthod Dentofac Orthop* 2014;145:1.
- [75] Barnes C, Boutron I, Giraudeau B, Porcher R, Altman DG, Ravaud P. Impact of an online writing aid tool for writing a randomized trial report: the COBWEB (Consort-based WEB tool) randomized controlled trial. *BMC Med* 2015;13:221.
- [76] Blanco D, Biggane AM, Cobo E. Are CONSORT checklists submitted by authors adequately reflecting what information is actually reported in published papers? *Trials* 2018;19:80.
- [77] Moher D, Altman DG. Four proposals to help improve the medical research literature. *PLoS Med* 2015;12:e1001864.
- [78] Moher D, Jones A, Lepage L. Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation. *JAMA* 2001;285:1992–5.
- [79] Gray R, Sullivan M, Altman DG, Gordon-Weeks AN. Adherence of trials of operative intervention to the CONSORT statement extension for nonpharmacological treatments: a comparative before and after study. *Ann R Coll Surg Engl* 2012;94(6):388–94.
- [80] Nagendran M, Harding D, Teo W, Camm C, Maruthappu M, McCulloch P, et al. Poor adherence of randomised trials in surgery to CONSORT guidelines for non-pharmacological treatments (NPT): a cross-sectional study. *BMJ Open* 2013;3(12):1–7.
- [81] Sarkis-Onofre R, Poletto-Neto V, Cenci MS, Pereira-Cenci T, Moher D. Impact of the CONSORT Statement endorsement in the completeness of reporting of randomized clinical trials in restorative dentistry. *J Dent* 2017;58:54–9.
- [82] Agha R, Cooper D, Muir G. The reporting quality of randomised controlled trials in surgery: a systematic review. *Int J Surg* 2007;5(6):413–22.
- [83] Savithra P, Nagesh LS. Have CONSORT guidelines improved the quality of reporting of randomised controlled trials published in public health dentistry journals? *Oral Health Prev Dent* 2013;11(2):95–103.
- [84] Macpherson H, Altman DG, Hammerschlag R, Youping L, Taixiang W. Revised STANDards for reporting interventions in clinical trials of acupuncture (STRICTA): extending the CONSORT statement. *PLoS Med* 2010;7(6):e1000261.
- [85] Asher GN, Motsinger-Reif AA, Jonas DE, Viera AJ. Quality of reporting on randomised controlled trials of auriculotherapy for pain. *Acupunct Med* 2011;29(2):122–6.
- [86] Ma B, Chen ZM, Xu JK, Wang YN, Chen KY, Ke FY, et al. Do the CONSORT and STRICTA checklists improve the reporting quality of acupuncture and Moxibustion randomized controlled trials published in Chinese journals? A systematic review and analysis of trends. *PLoS One* 2016;11:e0147244.
- [87] Svenkerud S, MacPherson H. The impact of STRICTA and CONSORT on reporting of randomised control trials of acupuncture: a systematic methodological evaluation. *Acupunct Med* 2018;36:349–57.
- [88] Kim KH, Kang JW, Lee MS, Lee J-D. Assessment of the quality of reporting in randomised controlled trials of acupuncture in the Korean literature using the CONSORT statement and STRICTA guidelines. *BMJ Open* 2014;4(7):e005068.
- [89] Zeng J, Lin G, Li L, Lu L, Chen C, Lu L. Assessment of reporting quality in randomised controlled trials of acupuncture for post-stroke rehabilitation using the CONSORT statement and STRICTA guidelines. *Acupunct Med* 2017;35(2):100–6.
- [90] Jia P, Tang L, Yu J, Liu J, Kang D, Sun X. The quality of reporting in randomized controlled trials of acupuncture for knee osteoarthritis: a cross-sectional survey. *PLoS One* 2018;13:e0195652.
- [91] Yamato TP, Maher CG, Saragiotto BT, Moseley AM, Hoffmann TC, Elkins MR. The TIDieR checklist will benefit the physiotherapy profession. *J Physiother* 2016;62(2):57–8.

4.2 Resultados del segundo trabajo

4.2.1 Descripción de los artículos

Los 120 ECAs analizados evaluaron el efecto de seis modalidades diferentes de terapias manuales (técnicas de tejidos blandos (29%), terapia de manipulación espinal (23%), movilización articular (21%), tratamientos quiroprácticos (12%), acupresión / reflexología (10 %) y tratamiento de manipulación osteopática (5%)) comparado con hasta 20 tipos diferentes de intervenciones de control. En algunos estudios se utilizó más de una intervención de control. La “atención habitual” (21%), las intervenciones simuladas (19%) y el ejercicio (18%) fueron las intervenciones de control más comunes. Los ensayos incluyeron una media de 111 participantes (SD=187). La mayoría de estudios evaluaron las respuestas fisiológicas en sujetos sanos o asintomáticos, y los problemas de salud más frecuentes evaluados fueron el dolor lumbar y el dolor cervical. El treinta y nueve por ciento de los artículos reportaron un seguimiento a largo plazo (>3 meses desde el fin de la intervención). Los artículos estudiados fueron publicados en revistas con un factor de impacto que oscilaba entre 0,12 y 4,63 en el *Scientific Journal Ranking* (Scimago). En el año de su publicación, el 74% eran revistas del primer cuartil, 20% del segundo cuartil y 6% del tercer cuartil.

4.2.2 Número de limitaciones

Inicialmente, se identificó una lista de 53 limitaciones específicas autoreportadas por los autores, y a partir de la discusión entre revisores para establecer una definición clara de cada limitación específica y después del agrupamiento temático, se llegó a una lista final de 39. Las limitaciones se encontraron ubicadas casi exclusivamente en la sección de discusión de los artículos. Solo dos artículos (1,6%) incluyeron una sección específica de limitaciones en el resumen. El número de LARs por artículo osciló entre 0 y 8, y el 80% de los artículos reconocieron al menos 2 limitaciones diferentes. En el 9% de los artículos no se encontraron LARs.

4.2.3 Categorías y limitaciones específicas reportadas

La limitación más comúnmente declarada (en casi la mitad de nuestra muestra) fue la relacionada con el tamaño de la muestra (47,5%), seguida de las limitaciones relacionadas con la duración del estudio y el seguimiento (33,3%) y controles inadecuados (32,5%). Las limitaciones respecto a la financiación del estudio o relacionadas con los centros seleccionados y / o proveedores de atención fueron las menos reportadas. También se analizaron las razones que los autores reportaron para justificar la inclusión de una determinada limitación. En algunos casos, los autores informaron de diferentes aspectos del estudio que conducían a una misma limitación específica. Por ejemplo, con respecto a las limitaciones del tamaño de la muestra, en el 87,7% de los artículos los autores declararon una falta de poder estadístico para proporcionar resultados fiables, mientras que solo el 8,8% declaró dificultades de reclutamiento. En estos casos, aunque se registraron los motivos específicos de cada situación, ambos artículos fueron incluidos bajo la misma categoría de estudios con limitaciones relacionadas con el tamaño de la muestra.

4.2.4 Segunda publicación

Alvarez Gerard, Núñez-Cortés Rodrigo, Solà Ivan, Sitjà-Rabert Mercè, Fort-Vanmeerhaeghe Azahara, Fernández Carles, Bonfill Xavier, Urrútia Gerard. ***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. 32 October 2020 . doi.org/10.1016/j.jclinepi.2020.10.018. IF (2020): 6.4

ORIGINAL ARTICLE

Sample size, study length, and inadequate controls were the most common self-acknowledged limitations in manual therapy trials: A methodological review

Gerard Alvarez^{a,b,*}, Rodrigo Núñez-Cortés^c, Ivan Solà^{a,d}, Mercè Sitjà-Rabert^{c,f}, Azahara Fort-Vanmeerhaeghe^{c,g}, Carles Fernández^{c,f}, Xavier Bonfill^{a,d}, Gerard Urrútia^{a,d}

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

^bFoundation Centre for Osteopathic Medicine Collaboration, Spain National Centre, Barcelona, Spain

^cDepartment of Physical Therapy, Faculty of Medicine, University of Chile, Santiago, Chile

^dCIBER Epidemiología y Salud Pública (CIBERESP), Spain

^eBlanquerna School of Health Science (FCS), Ramon Llull University, Barcelona, Spain

^fGlobal Research on Wellbeing (GRoW) Research Group, Ramon Llull University, Barcelona, Spain

^gBlanquerna Faculty of Psychology, Education Sciences and Sport (FPCEE), Ramon Llull University, Barcelona, Spain

Accepted 28 October 2020; Published online 1 November 2020

Abstract

Objectives: The aim of this study was to quantify and analyze the presence and type of self-acknowledged limitations (SALs) in a sample of manual therapy (MT) randomized controlled trials.

Study Design and Setting: We randomly selected 120 MT trials. We extracted data related to SALs from the original reports and classified them into 12 categories. After data extraction, specific limitations within each category were identified. A descriptive analysis was performed using frequencies and percentages for qualitative variables.

Results: The number of SALs per trial article ranged from 0 to 8, and more than two-thirds of trials acknowledged at least two different limitations. Despite its small proportion, 9% of trials did not report SALs. The most common limitation declared, in almost half of our sample, related to sample size (47.5%) followed by limitations related to study length and follow-up (33.3%) and inadequate controls (32.5%).

Conclusion: Our results indicate that at least two different limitations are consistently acknowledged in MT trial reports, the most common being those related to sample size, study length, follow-up, and inadequate controls. Analysis of the reasons behind the SALs gives some insights about the main difficulties in conducting research in this field and may help develop strategies to improve future research. © 2020 Elsevier Inc. All rights reserved.

Keywords: Self-acknowledged limitations; Manual therapy; Reporting; Quality; Transparency; Guidelines

Author statement: G.A. conceptualized and designed the study. G.A. and G.U. were responsible to create the first version of the data extraction form and G.A., I.S., M.S., A.F., C.F., and G.U. piloted the form and contributed to reach the final version. I.S. was responsible of the search strategy and articles retrieval. I.S., M.S., A.F., C.F., and G.U. were involved in the data extraction process forming pairs with G.A., R.N. and G.A. were responsible of data analysis. G.A. provided the first manuscript draft which received critical revision by G.U., I.S., and X.B.. All authors read and approved the final manuscript. **Gerard Alvarez is a PhD student on Biomedical Research Methodology and Public Health in the Medical Department of the Universitat Autònoma de Barcelona, Barcelona, Spain.**

Funding sources: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of interest: The authors declare that they have no competing interests.

* Corresponding author. Iberoamerican Cochrane Center C, Sant Antoni Maria Claret 167 Pavelló 18, Planta 0 08025 Barcelona, Catalunya, Spain. Tel.: +34 620 214 206.

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

<https://doi.org/10.1016/j.jclinepi.2020.10.018>

0895-4356/© 2020 Elsevier Inc. All rights reserved.

Key findings

- Over two-thirds of manual therapy trials acknowledged at least two different limitations.
- They are mostly related to sample size, study length and follow-up, and inadequate controls.
- In the 9% of the trials, no limitations are acknowledged.

What this adds to what was known?

- Identifying the most self-acknowledged limitations in a large sample of manual therapy trials provides a better understanding of research-related difficulties in that field.

What is the implication and what should change now?

- Future trials in the field of manual therapy should be designed in a way that these main limitations are overcome.
- We recommend the development of specific guidelines on how to report and acknowledge limitations to enhance research quality and transparency.

1. Introduction

Limitations of a study are features of its design and execution that could have affected the interpretation, validity, and applicability of the findings [1]. As they can never be excluded, authors should always consider acknowledging the potential limitations of their work [2]. In fact, recognition and discussion of potentially important limitations and their likely implications on the interpretation of the findings represent a central part of the scientific discourse. Moreover, such discussion can benefit the scientific community as it allows the investigators of future studies to address these limitations and produce higher quality studies [3].

This notwithstanding, insufficient acknowledgment of limitations in discussion sections has been previously found to be among the most important deficiencies in scientific articles [4]. Some studies have reported a lack of or inappropriate recognition of limitations in the medical literature. A study by Ioannidis, using automated keyword searching, found that only 17% of 400 articles published in leading medical journals used at least one word referring to limitations [5]. Ter Riet et al. analyzed the reporting of limitations and the use of “hedging patterns” (i.e., modal verbs like “may” or “could,” adverbs like “apparently” or “possibly,” and lexical verbs like “suggest [that]” or “believe [that]”) in 300 biomedical

publications published in 30 high- and medium-ranked journals in 2007. They found that 27% of the articles did not report any limitations, and that there were major differences between journals in how uncertainty was expressed and limitations acknowledged. Despite improvement over time, the conclusion was that the reporting of limitations was probably incomplete [6]. Collectively, this literature shows that failing to acknowledge limitations is still a significant problem in the biomedical literature. Proposed strategies to help improve transparency and contribute to the evolving culture of research reproducibility include biomedical text mining to automatically recognize self-acknowledged limitations (SALs) in clinical research publications and “second independent” discussions (the participation of an additional author who is independent to the study and whose role is to identify its limitations) [7–9].

Limitations may be omitted for several reasons. Researchers may feel under pressure to show that their work is important, robust, and limitation-free [5]. Acknowledging limitations may be perceived by authors as a source of negative comments from peer reviewers or even as a motive for nonacceptance by journals [3]. However, there is a compelling argument that articles that fully attempt to address limitations are likely to be widely cited and shape the future research agenda [3,9]. Manuscript length limitations in traditional print journals may be another reason to avoid or be sparse in acknowledging limitations. In fact, the superiority of open-access journals in acknowledging limitations may reflect space availability [5]. The increasing existence of open-access journals and online publishing should be seen as a solution regarding length restrictions.

Manual therapy (MT) is a physical treatment used by a variety of therapists to treat mainly musculoskeletal pain and disability [10]. It includes massage, joint mobilization/manipulation, myofascial release, nerve manipulation, strain/counterstrain, and acupressure [11]. In the field of physical and MT, although the number of publications has increased significantly in recent years, this does not seem to have been accompanied by improved quality or reporting [12–16]. Because of its nature, research in complex, nonpharmacological interventions is exposed to a number of methodological difficulties that can affect both internal and external validity [17,18], leading to various study limitations [19]. These weaknesses have been previously reported in several publications focused on the quality and reporting of nonpharmacological randomized controlled trials (RCTs) [20–22]. Given this situation, and for the sake of a correct interpretation of the findings, a complete description of SALs should be expected in this kind of literature. However, to the best of our knowledge, no attempts have been made to evaluate this aspect in MT articles.

The aim of this study was to quantify and analyze the presence and type of SALs in a sample of MT RCTs.

2. Methods

2.1. Eligibility criteria

We report a secondary analysis from a previous systematic methodological evaluation of a random sample of 100 MT RCTs published between 2000 and 2015 [15]. While we first focused on reporting, the present study assessed the presence and type of SALs in the sample. For the purposes of the present study, we extended our sample with 20 additional trials published from 2015 to 2020 to assure the validity of the results and to be representative of articles published up to the present day. Criteria for inclusion were that at least one of the interventions (experimental or control) should include some form of MT. A validated search strategy (Appendix 1) was used to retrieve MT articles [23]. More details on eligibility criteria, search strategy, and a description of the reviewers are available elsewhere [15].

2.2. Data extraction

Two researchers, one with expertise in clinical research methodology (GU) and the other one with expertise in the field of MT (GA), defined by consensus, a list of 12 items indicating broad categories of common limitations in conducting clinical trials. We developed a data extraction form including these categories and an additional open field to register any other limitation beyond those defined.

The reviewers (GA, IS, MS, AF, CF, and GU) were first asked to identify any of the limitations defined in the abstract and discussion sections from the trial reports and then to transcribe the information related to the SALs identified. We extracted data in pairs of independent reviewers.

2.3. Data analysis

From the data extracted, we accounted for the trials disclosing the predefined limitations. Two reviewers (GA and RN) compared iteratively the limitations descriptions reported in the trial reports to obtain common themes and define specific limitations implementing a thematic analysis [24]. Descriptive analyses were performed using SPSS version 22.0 (IBM Corporation, Armonk, NY), using frequencies and percentages of qualitative variables, including the 95% confidence interval.

3. Results

3.1. Description of articles

The 120 trials analyzed assessed the impact of six different modalities of manual therapies (soft tissue techniques (29%), spinal manipulation therapy (23%), joint mobilization (21%), chiropractic treatments (12%), acupressure/reflexology (10%), and osteopathic manipulative treatment (5%)) comparing them with up to 20

different types of control treatments/interventions. In some studies, more than one control intervention was used, and “usual care” (21%), sham interventions (19%), and exercise (18%) were the most common. The trials included a mean of 111 participants (SD = 187). The most frequent studies evaluated physiological responses on healthy or asymptomatic subjects, and the most frequent health problems assessed were lower back pain and neck pain. Thirty-nine percent of the articles reported long-term follow-up. We reported a full description of the articles included elsewhere [15].

The articles studied were published in journals that ranged from 0.12 to 4.63 in the scientific journal ranking (Scimago). At the year of publication, 74% of them were first quartile journals, 20% were second quartile, and 6% were third quartile [25].

3.2. Number of limitations

Initially, a list of 53 specific limitations was identified, and after discussion between reviewers to establish a clear definition of each specific limitation, a final list of 39 was reached (Table 1). Limitations were located almost exclusively in the discussion section. Only two articles (1.6%) included a specific limitation section in the Abstract. The number of SALs per article ranged from 0 to 8, and 80% of the articles acknowledged at least 2 different limitations. In 9% of the articles, no SALs were found (Fig. 1).

3.3. Categories and specific limitations reported

Table 2 shows the frequency and categories of SALs reported in the articles. The most common limitation declared, in almost half of our sample, related to sample size (47.5%), followed by limitations related to study length and follow-up (33.3%) and inadequate controls (32.5%). Limitations regarding funding or the selected centers and/or care providers were the least reported. For each category of limitations, several explanations were given by the authors, and in some cases, they reported different study weaknesses leading to the same type of specific limitation. Table 2 shows the frequency of these explanations, which allow better understanding of the specific limitations within each category. For example, regarding sample size limitations, in 87.7% of the articles, authors declared a lack of statistical power to provide reliable results, while just 8.8% declared recruitment difficulties.

4. Discussion

In the present study, we quantified and analyzed SALs in a sample of 120 manual therapy trials. While previous research assessing SALs has focused on high-impact journals regardless the study field [5,6], we extended our focus to trials published in any journal to capture the breadth of this kind of clinical research in manual therapy. Overall,

Table 1. Type of limitations and reasons reported by authors

Category 1—randomization and/or concealment	
Unbalanced groups	After randomization, the number of subjects allocated to each group was unequal.
Lack of stratification	Randomization method did not include stratification.
Selection bias not excluded	Randomization methods were not the most recommended (e.g., ballot boxes) or allocation was not concealed.
Category 2—Blinding	
Statistician blinding	People performing the statistical analysis were not blinded.
Outcome assessor blinding	People in charge of data collection were not blinded.
Patient blinding	Patients were not blinded with respect to the study groups.
Care provider blinding	People providing the interventions were not blinded with respect to the study groups.
Category 3—Lost to follow-up/dropouts	
Unbalanced dropouts	Dropout rate between study groups was unequal.
Many lost to follow-up	Discontinuation rate was superior than expected/estimated leading to underpowered studies and less precise results.
Category 4—Sample size	
Unpowered study	Inability to detect differences between groups due to sample size.
Convenience sampling	The sampling method was linked to specific study needs.
Recruitment less than expected	The sample size did not reach the estimated number due to recruitment difficulties.
Category 5—Inadequate control	
Active control	An active intervention (e.g., exercise or usual care) was selected as control.
Simulation (sham) not optimal	Some effects related to sham interventions could not be ruled out.
Different parameters between groups	The dosage, the number of sessions, or their duration was different in the control group.
Different baseline parameters.	Comparison of baseline data showed significant differences in the control group
Category 6—Funding	
Limited funding	The limited or lack of funding affected the study progress or completion.
Category 7—Type of setting/care provider/s	
Unicentric	Study was conducted recruiting participants from a single center.
Number of care providers	Intervention delivered by a single investigator.
Category 8—Inadequate subjects	
Asymptomatic subjects	Leading to a limited clinical interpretation of the results.
Diagnostic criteria	Lack of standardized diagnostic criteria to include participants.
Very specific population	Inclusion criteria considered too restricted. (e.g., single gender, athletes only, education level, or race).
Category 9—Study length and follow-up	
No long-term follow-up	Only short-term effect was evaluated.
Missing data due to study length	Dropouts or losses to follow-up derived from a very long follow-up time.
Unequal follow-up	Follow-up between the study groups was different.
Category 10—Intervention	
Additive effects	It was not possible to know the net effect of every component in multimodal treatments.
Limited treatment area	The intervention was applied to a very specific body region (e.g., single joint or vertebral segment).
Treatment parameters not standardized	Specific parameters of manual techniques were not standardized (e.g., amount of force, direction, or speed).
Lack of individualization	A treatment protocol was applied without adapting the interventions to the characteristics of the individuals or to the diagnostic criteria.
Therapist profile	Care providers required specific training on the intervention, or care providers had different previous experience

(Continued)

Table 1. Continued

Category 11—Outcome measures	
Reliability of the measurement	Lack or low reliability of outcome measures (e.g., self-reported results).
Lack of measurement data	Some relevant data were not collected during the study (e.g., demographics or clinical variables that would potentially provide interesting findings).
Measurement sensitivity	Outcome measures were not sensitive enough to detect subtle changes (e.g., use of ordinal scales).
Validity of the measurement	The selected assessment instrument was not originally validated for that specific population.
Suboptimal statistical analysis	Findings were not adjusted for covariates.
Category 12—Compromised generalization of the results (This category includes those self-acknowledged limitations that were mentioned specifically in regards to their impact on the generalization of the results or external validity.)	
Compromised generalization due to type of setting.	The characteristics of the setting where the study was conducted limited the generalization of the results
Compromised generalization due to specific population	The subjects' characteristics of the sample limited the generalization of the results.
Compromised generalization due to measurement	The measurement instrument selected to assess the effect of the intervention sample limited the generalization of the results.
Compromised generalization due to intervention	The type of intervention or the way it was delivered limited the generalization of the results.

our results indicate that at least two different limitations are consistently acknowledged in MT reports. Furthermore, the analysis of the arguments reported by authors provide some explanation to help understand the specific difficulties that MT investigators face and in turn establish strategies, recommendations, or guidelines to improve these. The following discussion focuses on those SALs present in at least 20% of our sample.

4.1. Limitations related to sample size

The most reported SAL related to sample size (47.5%) and, within this category, the main reason given by the authors was that the studies lacked statistical power to detect effect differences between groups (87.7%). In general, RCTs in MT frequently involve small sample sizes [26]. Research in MT is often conducted in private centers and outpatient services which limits recruitment possibilities, leading to underpowered studies. For example, in our sample, although

pilot studies were not excluded ($n = 15$), only 31% of the RCTs had samples greater than 100 participants [15].

It has been found that sample size is related to an a priori sample size calculation: trials that perform an a priori sample size calculation have considerably larger median sample sizes than those that do not [27]. However, despite improvement over time, reporting of sample size calculation and power analysis remains inadequate regardless the study type [28]. In the field of rehabilitation trials, in a sample of 222 RCTs, Castellini et al. found that only 36% reported sample size calculations [29]. In the same line, Gonzalez et al. concluded that just 34% of the physical therapy trials relevant to musculoskeletal conditions reported adequate sample size calculation [14]. In our sample, a description of sample size calculation was found in 49% of the articles. Therefore, a statement of sample size calculation should be seen as a key requirement for MT investigators.

While adequate sample size is desirable, some authors have argued that underpowered trials might be acceptable if investigators use methodological rigor to eliminate bias, report properly to avoid misinterpretation, and always publish results to avoid publication bias [30]. According to Schulz et al., readers should be more concerned about systematic errors rather than inadequate sample size. In fact, sample size deals more with the precision of the estimation of effect and not necessarily with the validity of the study [26]. The power of the trial is expressed in the confidence interval. Hence, if methodological issues are carefully considered, the power is no longer a major concern [30]. In turn, validity can dramatically influence the effect size.

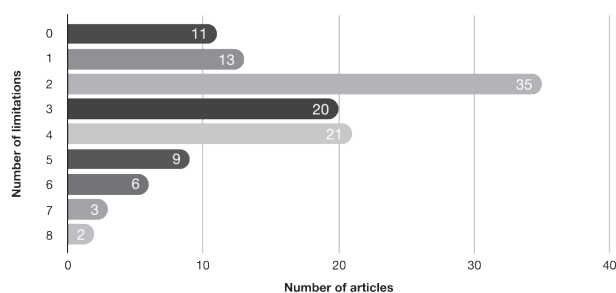


Fig. 1. Number of limitations reported in the articles ($n = 120$).

For example, in the field of low back pain, studies with higher risk of bias reported effect sizes that were, on average, 50% greater than the estimates reported in trials with a lower risk of bias [31]. In a sample of trials evaluating treatments for hip and knee osteoarthritis, it has been shown that inadequate random sequence generation and lack of allocation concealment and double-blinding yielded larger treatment effects [20].

4.2. Limitations related to inadequate control intervention

In the field of physical, manual, and rehabilitation medicine, the lack of high-quality clinical research can be partially explained by the inherent difficulty in conducting double-blinded placebo-controlled clinical trials [32,33]. In our sample, limitations related to inadequate control intervention were reported in 32.5% of the reviewed articles. Within this category, the main reason stated by authors (41%) pointed to placebo or sham interventions not being optimal for the comparison. In the field of MT (or any discipline with a high interaction between patients and care providers), the use of placebos or sham interventions is highly influenced by the therapist who delivers the treatment [26,33]. Thus, in manual sham RCTs, isolating the “active ingredient” from other effects can be very complex [34]. However, key authors in the field of MT conceptualize placebo as an active and important mechanism in MTs [35], including the placebo response as an inseparable and “always present” mechanism of MT effect [36,37].

Therefore, recommendations for future research should include both improvements in placebo/sham procedures and strategies to better assess and report the placebo response in both the intervention and the control group. By using appropriate placebos, it is possible to minimize several kinds of bias in RCTs and to improve research quality. However, there is still a lack of a common, valid manual placebo gold-standard [34]. Moreover, although methodological recommendations have been made by authors in the field of rehabilitation and MTs [32,38,39] and the Consolidated Standards of Reporting Trials (CONSORT) and other statements provide guidelines to report control interventions, there are no clear, unified guidelines on the design or description of sham/placebo procedures [21,40]. Within this scenario, we emphasize once more the importance of complete reporting. For example, in a review of 64 studies in the field of osteopathy, Cerritelli et al. found that trials systematically underreported details on sham procedures, sham dosage, and sham operator. These findings, alongside the high heterogeneity and within-study variability between sham and real treatment procedures profoundly compromise study validity [34]. Fortunately, an extension of the Template for Intervention Description and Replication (TIDieR) for placebo-controlled trials (TIDieR–Placebo) is currently registered in the EQUATOR Network as a guideline under development [41], which is

expected to be a useful tool to facilitate the design of better research in the field.

4.3. Limitations related to blinding

Control-related limitations are linked to blinding strategies, which, in our sample, accounted for 28.3% of all SALs (Table 2). In fact, when comparing clinical trials evaluating nonpharmacological vs. pharmacological treatments, lack of blinding in nonpharmacological treatment articles explained most of the methodological differences [20]. In the field of physical medicine and rehabilitation, reporting of blinding still does not follow current recommendations [42]. Despite the recent controversies raised on the role of the impact of blinding on effect estimates in randomized trials [43], in that particular field, lower reporting of blinding among trials with positive outcomes have been reported [42]. Blinding is less frequently reported in RCTs assessing nonpharmacological interventions, possibly due to the difficulty in achieving and maintaining it [21]. While blinding outcome assessor(s) should be feasible in most cases, blinding patients is challenging and almost impossible for manual therapists. Addressing this issue specifically, Boutron et al. published in 2007 a list of blinding methods used in nonpharmacological trials and provided guidelines to overcome some related difficulties [21]. Among other possible strategies, they proposed blinding participants to study hypotheses (modified Zelen design) and the use of attention-control interventions or placebo control interventions that are not identical to the active treatment [21]. The strengths and limitations of each method should be carefully considered in every case. In the other hand, trials with subjective outcomes, complex or operator-dependent interventions such as MTs could more easily be biased by beliefs and expectations [44]. The strength and direction of preconceived beliefs likely modifies the effect of blinding [44], and contextual factors effects are well documented in MT literature [45–47]. In this scenario, measuring credibility of the treatment, participants’ expectations or registering participants’ preferences at the beginning of the trial before randomization could help to assess the strength of the treatment [21]. Finally, blinding is rarely tested, and there is methodological uncertainty on how to assess it [48]. Although a test of blinding success might be seen as a good standard practice to adopt, difficulties to assess its impact on potential bias have been reported [49,50]. On the contrary, pretrial assessment of blinding procedures and explicit reporting of who was blinded and by what means is strongly recommended [51].

4.4. Limitations related to study length and follow-up

Study length and follow-up is another common limitation in MT trials (33.3% in our sample). The main study length limitations declared related to the lack of a long-term follow-up (87.5%). Collectively, the literature on specific

Table 2. Frequency and type of reported limitations

C1	Randomization and/or concealment	N articles	Frequency [CI 95%]
<i>N</i> = 13/10.8% [5.9 to 17.8]	Unbalanced groups	2	15.4% [1.9 to 45.4]
	Lack of stratification	3	23.1% [5.1 to 53.8]
	Selection bias not excluded	8	61.5% [31.5 to 86.1]
C2	Blinding	N articles	Frequency [CI 95%]
<i>N</i> = 34/28.3% [20.5 to 37.3]	Statistician blinding	1	2.9% [0.1 to 15.3]
	Outcome assessor blinding	4	11.8% [3.3 to 27.5]
	Patient blinding	24	70.6% [52.5 to 84.9]
	Care provider blinding	17	50.0% [32.4 to 67.6]
C3	Lost to follow-up/dropouts	N articles	Frequency [CI 95%]
<i>N</i> = 18/15% [9.1 to 22.7]	Unbalanced dropouts	1	5.6% [0.1 to 2.67]
	Many lost to follow-up	17	94.4% [72.7 to 99.9]
C4	Sample size	N articles	Frequency [CI 95%]
<i>N</i> = 57/47.5% [38.3 to 56.8]	Unpowered study	50	87.7% [76.3 to 94.9]
	Convenience sampling	2	3.5% [0.4 to 12.1]
	Recruitment less than expected	5	8.8 [2.9 to 19.3]
C5	Inadequate control	N articles	Frequency [CI 95%]
<i>N</i> = 39/32.5% [24.2 to 41.7]	Active control.	14	35.9% [21.2 to 52.8]
	Simulation (sham) not optimal	16	41% [25.5 to 57.9]
	Different parameters between groups	4	10.3% [2.9 to 24.2]
	Different baseline parameters	5	12.8% [4.3 to 27.4]
C6	Funding	N articles	Frequency [CI 95%]
<i>N</i> = 5/4.2% [1.4 to 9.5]	Limited funding	5	100% [47.8 to 100]
C7	Type of setting/care providers	N articles	Frequency [CI 95%]
<i>N</i> = 9/7.5% [3.5 to 13.8]	Unicentric	8	88.9% [51.7 to 99.7]
	Number of care providers	1	11.1% [0.3 to 48.2]
C8	Inadequate subjects	N articles	Frequency [CI 95%]
<i>N</i> = 31/25.8% [18.3 to 34.6]	Asymptomatic subjects	4	12.9% [3.6 to 29.8]
	Diagnostic criteria	5	16.1% [5.4 to 33.7]
	Very specific population	23	74.2% [55.4 to 88.1]
C9	Study length and follow-up	N articles	Frequency [CI 95%]
<i>N</i> = 40/33.3% [25.0 to 42.5]	No long-term follow-up	35	87.5% [73.1 to 95.8]
	Missing data due to study length	2	5% [0.6 to 16.9]
	Unequal follow-up	3	7.5% [1.5 to 20.4]
C10	Intervention	N articles	Frequency [CI 95%]
<i>N</i> = 33/27.5% [19.7 to 36.4]	Additive effects	6	18.2% [6.9 to 35.5]
	Limited treatment area	8	24.2% [11.1 to 42.3]
	Treatment parameters not standardized	17	51.5% [33.5 to 69.2]
	Lack of individualization	5	15.2% [5.1 to 31.9]
	Therapist profile	1	3% [0.1 to 15.8]
C11	Outcome measures	N articles	Frequency [CI 95%]
<i>N</i> = 35/29.2% [21.2 to 38.2]	Reliability of the measurement	9	25.7% [12.5 to 43.3]
	Lack of measurement data	12	34.3% [19.1 to 52.2]
	Measurement sensitivity	2	5.7% [0.7 to 19.2]
	Validity of the measurement	11	31.4% [16.9 to 49.3]
	Suboptimal statistical analysis	4	11.4% [3.2 to 26.7]
C12	Compromised generalization (CG) of the results (<i>n</i> = 26)	N articles	Frequency [CI 95%]
<i>N</i> = 26/21.7% [14.7 to 30.1]	CG due to type of setting	3	11.5% [2.4 to 30.2]
	CG due to specific population	16	61.5% [40.5 to 79.8]
	CG due to measurement	2	7.7% [0.9 to 25.1]
	CG due to intervention	5	19.2% [6.5 to 39.4]

95% confidence interval (CI).

Note: Some articles reported more than one reason (specific limitations) within the same category.

mechanisms of MT concludes that the main effects are short term [36,37]. However, in clinical practice, MT is commonly used to treat chronic conditions [45,52] through therapeutic packages that can last several weeks or months. Therefore, long-term follow-up should be encouraged in MT clinical trials assessing chronic conditions. Ideally, MT investigators should develop intervention models that match the natural history of common problems, considering the trajectory of the symptoms over time as an outcome rather than focusing on specified follow-up time points alone [53].

4.5. Limitations related to inadequate subject selection

Inadequate subject selection appeared as a SAL in 25.8% of our sample and can be linked to limited generalization (21.7%). In both cases, the main reason behind these limitations was the recruitment of a very specific population (Table 2). In a considerable number of cases, asymptomatic subjects were recruited to evaluate the effect of manual interventions on physiological parameters, which may differ in a broader population of patients in a more real-world setting. In other cases, strict inclusion criteria were established, limiting the generalization of the results. As complex, nonpharmacological, person-centered interventions, MTs are delivered as part of a care package rather than as a single treatment [54]. In addition, at least in the field of musculoskeletal problems, patients typically have multiple comorbidities [55–57]. Therefore, a pragmatic attitude should be adopted in MT trials, to try to include little or no participant selection beyond the clinical indication of interest [58]. However, to reach internal validity standards and allow replication, MT trialists might be forced to highly select the sample, compromising the generalization of the results.

4.6. Limitations related to the intervention

Limitations related to the intervention were acknowledged in 27.5% of our sample. The main reason given by the authors was the lack of a standardized protocol (51.5%), which contrasts with the lack of individualization also reported as a specific limitation in 15.2% of this category. Furthermore, whether the treatment has to be applied in a certain localized body region or with a whole-body perspective can also be problematic. In our sample, 24.2% of intervention-related SALs were explained by the treatment being delivered to only a certain area of the body. This reflects again the struggle that MT researchers have in finding the balance between real clinical practice and high-standard methodological requirements or, in other words, between reproducibility and clinical replicability [59]. This aspect is still an open debate that hinders the design and reporting of some interventions. The reporting of nonpharmacological interventions is poor and incomplete [22], and that includes all fields of physical therapy interventions [60] and specifically MT [13,15]. However, beyond the reporting, there is a need to recognize the

particularities of patient-centered touch-based interventions and strike a balance between internal and external validity. Some of the most recent reporting guidelines (e.g., the updated CONSORT extension for nonpharmacological interventions (CONOSRTnpt) [61] or the TIDieR) [22] already include items related to the tailoring of interventions. However, recent research shows that despite the existence of such guidelines, clinical reproducibility is compromised particularly for complex interventions and, as we proposed in a previous publication [15], there is a need for field-specific checklists [59]. Examples of this are the specific reporting guidelines on spinal manipulative therapy [62] and therapeutic exercise [63].

4.7. Limitations related to outcome measures

Finally, SALs related to outcome measures accounted for 29.2% of the limitations reported in our sample. Among the reasons given by the authors, the lack of measurement data and the validity of measurements were the most frequent (34.3% and 31.4%). The selection of valid and meaningful outcome measures is crucial to make research informative to clinical practice, and the outcome measurement tools used in research may not always be the most appropriate in clinical settings [64]. Furthermore, difficulties caused by heterogeneity in outcome measurements are constantly reported in systematic reviews [65] and constitute a challenge in meta-analysis of physical therapy interventions [66]. The use of a standardized set of outcome measures for a given condition (which include different types of outcomes) would be a potential solution to these issues. The Core Outcome Measures in Effectiveness Trials initiative (<http://www.comet-initiative.org>) represents a very useful tool for MT investigators. As an additional recommendation, the use of patient-reported outcomes measures (PROMs) would be a suitable option for patient-centered interventions such as MT. In fact, PROMs are increasingly collected in physiotherapy research and can also be used in clinical settings [67]. A specific CONSORT extension focused on PROMs [68] provides guidelines to researchers to ensure optimal reporting.

4.8. Comment on reported limitations

In the previous sections, we have quantified and described those SALs identified in our sample. However, beyond the descriptive purposes of the present study, this issue deserves further comment to discuss SALs nature, their implications, and impact. Although reported in a different proportion, we could differentiate some SALs with a greater impact on the internal trial validity from others not directly linked to the study design or that reflected issues related to its conducting. Within the first group, eight trials acknowledged potential selection bias due to inconsistencies in the randomization/concealment process, three trials acknowledged unequal follow-up between groups, and four acknowledged suboptimal statistical

analysis. In the second group, we identified SALs that were not truly methodological (e.g., limited funding in five trials) or reported problems derived from the trial progress (e.g., unbalanced groups in two trials) which should be amendable with a rigorous and accurate data analysis.

In accordance with its nature, SALs can have an impact either on the confidence of the trial's effect estimates or to its conclusions, and this could have some implications to practice. Researchers should ensure the acknowledgment of at least those limitations that might entail threats to the study validity when preparing the manuscripts to report the findings of the trial. On the other hand, reporting guidelines should include more specific guidance to improve the accuracy in reporting limitations, defining some mandatory or those highly recommended to report. Journals also should facilitate tools to appraise this issue and its implications during the peer review process.

4.9. Study limitations

We acknowledge several limitations in our study. First, although we selected 20 additional trials to update our original sample with RCTs published between 2015 and 2020, most recent articles might be slightly underrepresented. Second, the categorization of the 12 broad SALs and the list of specific limitations within each category was subjectively established by three members of the research team (GA, GU, and RN). To minimize potential bias in this categorization and to have common criteria for the allocation of every specific limitation to its category, several meetings took place at the beginning of and during the process. Discrepancies were solved through discussion.

5. Conclusion

Recognition and discussion of all potentially important limitations of a study represent a crucial part of the scientific discourse. Our results indicate that at least two different limitations are consistently acknowledged in MT studies, and that those related to sample size, study length and follow-up, and inadequate controls were the most commonly reported. Besides the items already included in reporting guidelines, specific extensions on how to report and acknowledge limitations could potentially enhance research quality and transparency. Furthermore, researchers should be encouraged to report those limitations that could entail a threat to their studies validity. Systematic analysis of the reasons behind the SALs provides some insights on the main difficulties in conducting studies in this field and may prove helpful in the development of strategies to improve future research.

Acknowledgments

The authors wish to acknowledge Ms. Victoria Leo for her contribution in improving the manuscript.

Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jclinepi.2020.10.018>.

References

- [1] Price Editor JH, Murnan Assistant J. Research limitations and the necessity of reporting them. *Am J Heal Educ* 2013;35(2):66–7.
- [2] Knottnerus JA, Tugwell P. Improving the clinical significance of medical research. *J Clin Epidemiol* 2014;67:607–8.
- [3] Puhon MA, Akl EA, Bryant D, Xie F, Apolone G, ter Riet G. Discussing study limitations in reports of biomedical studies- the need for more transparency. *Health Qual Life Outcomes* 2012;10(1):23.
- [4] Goodman SN, Berlin J, Fletcher SW, Fletcher RH. Manuscript quality before and after peer review and editing at *Annals of Internal Medicine*. *Ann Intern Med* 1994;121:11–21.
- [5] Ioannidis JPA. Limitations are not properly acknowledged in the scientific literature. *J Clin Epidemiol* 2007;60:324–9.
- [6] ter Riet G, Chesley P, Gross AG, Siebeling L, Muggensturm P, Heller N, et al. All that glitters isn't gold: a survey on acknowledgment of limitations in biomedical studies. *PLoS One* 2013;8:e73623–6.
- [7] Kilicoglu H. Biomedical text mining for research rigor and integrity: tasks, challenges, directions. *Brief Bioinformatics* 2017;533(9900): 415–52.
- [8] Kilicoglu H, Rosemblat G, Malički M, ter Riet G. Automatic recognition of self-acknowledged limitations in clinical research literature. *J Am Med Inform Assoc* 2018;25(7):855–61.
- [9] Avidan MS, Ioannidis JPA, Mashour GA. Independent discussion sections for improving inferential reproducibility in published research. *Br J Anaesth* 2019;122:413–20.
- [10] French HP, Brennan A, White B, Cusack T. Manual therapy for osteoarthritis of the hip or knee - a systematic review. *Man Ther* 2011; 16(2):109–17.
- [11] Smith AR. Manual therapy: the historical, current, and future role in the treatment of pain. *ScientificWorldJournal* 2007;7:109–20.
- [12] 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(1):36–43.
- [13] 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(5):241–52.
- [14] Gonzalez GZ, Moseley AM, Maher CG, Nascimento DP. Methodologic quality and statistical reporting of physical therapy randomized controlled trials relevant to musculoskeletal conditions. *Arch Phys Med Rehabil* 2018;99:129–36.
- [15] Alvarez G, Sola I, Sitjà-Rabert M. A methodological review revealed that reporting of trials in manual therapy has not improved over time. *J Clin Epidemiol* 2020.
- [16] 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.
- [17] 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.
- [18] Clark AM. What are the components of complex interventions in healthcare? Theorizing approaches to parts, powers and the whole intervention. *Soc Sci Med* 2013;93:185–93.
- [19] Cook CE, Thigpen CA. Five good reasons to be disappointed with randomized trials. *J Man Manip Ther* 2019;27(2):63–5.
- [20] Boutron I, Tubach F, Giraudeau B, Ravaud P. Methodological differences in clinical trials evaluating nonpharmacological and pharmacological treatments of hip and knee osteoarthritis. *JAMA* 2003;290: 1062–70.

- [21] Boutron I, Guittet L, Estellat C, Moher D, Hróbjartsson A, Ravaud P. Reporting methods of blinding in randomized trials assessing non-pharmacological treatments. *PLoS Med* 2007;4(2):0370.
- [22] 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.
- [23] Pillastrini P, Vanti C, Curti S, Mattioli S, Ferrari S, Violante FS, et al. Using PubMed search strings for efficient retrieval of manual therapy research literature. *J Manipulative Physiol Ther* 2015.
- [24] Barnett-Page E, Thomas J. Methods for the synthesis of qualitative research: a critical review. *BMC Med Res Methodol* 2009.
- [25] Guerrero-Bote VP, Moya-Aneón F. A further step forward in measuring journals' scientific prestige: the SJR2 indicator. *J Informetr* 2012;6(4):674–88.
- [26] Koes BW. How to evaluate manual therapy: value and pitfalls of randomized clinical trials. *Man Ther* 2004;9(4):183–4.
- [27] Rubinstein SM, Van Eekelen R, Oosterhuis T, De Boer MR, Ostelo RW, Van Tulder MW. The risk of bias and sample size of trials of spinal manipulative therapy for low back and neck pain: analysis and recommendations. *J Manipulative Physiol Ther* 2014;37(8):523–41.
- [28] Lee PH, Tse ACY. The quality of the reported sample size calculations in randomized controlled trials indexed in PubMed. *Eur J Intern Med* 2017;40:16–21.
- [29] Castellini G, Gianola S, Bonovas S, Moja L. Improving power and sample size calculation in rehabilitation trial reports: a methodological assessment. *Arch Phys Med Rehabil* 2016;97:1195–201.
- [30] Schulz KF, Grimes DA. Sample size calculations in randomised trials: mandatory and mystical. *Lancet* 2005;365:1348–53.
- [31] Van Tulder MW, Suttrop M, Morton S, Bouter LM, Shekelle P. Empirical evidence of an association between internal validity and effect size in randomized controlled trials of low-back pain. *Spine (Phila Pa 1976)* 2009;34(16):1685–92.
- [32] 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.
- [33] Chaibi A, Šaltytė Benth J, Bjørn Russell M. Validation of placebo in a manual therapy randomized controlled trial. *Sci Rep* 2015;5(1):11774.
- [34] Cerritelli F, Verzella M, Cicchitti L, D'Alessandro G, Vanacore N. The paradox of sham therapy and placebo effect in osteopathy A systematic review. *Med (United States)* 2016;95(35).
- [35] Bialosky JE, Bishop MD, Penza CW. Placebo mechanisms of manual therapy: a sheep in wolf's clothing? *J Orthop Sports Phys Ther* 2017;47(5):301–4.
- [36] Bialosky JE, Bishop MD, Price DD, Robinson ME, George SZ. The mechanisms of manual therapy in the treatment of musculoskeletal pain: a comprehensive model. *Man Ther* 2009.
- [37] 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.
- [38] Hancock MJ, Maher CG, Latimer J, McAuley JH. Selecting an appropriate placebo for a trial of spinal manipulative therapy. *Aust J Physiother* 2006;52(2):135–8.
- [39] Hart T, Bagiella E. Design and implementation of clinical trials in rehabilitation research. *Arch Phys Med Rehabil* 2012;93:S117–26.
- [40] Benedetti F. Placebo effects: from the neurobiological paradigm to translational implications. *Neuron* 2014;84(3):623–37.
- [41] Howick J, Bishop F, Collins G. TIDieR—Placebo checklist explanation and elaboration. TIDieR—Placebo checklist explanation and elaboration. Available at: <https://www.equator-network.org/library/reporting-guidelines-under-development/reporting-guidelines-under-development-for-clinical-trials/#TIDIERPLA>. Accessed February 4, 2020.
- [42] 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(1):6–13.
- [43] 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;368:16802–13.
- [44] Drucker AM, Chan A-W. Blindsided: challenging the dogma of masking in clinical trials. *BMJ* 2020;368:m229.
- [45] Coronado RA, Bialosky JE. Manual physical therapy for chronic pain: the complex whole is greater than the sum of its parts. *J Man Manip Ther* 2017;25(3):115–7.
- [46] 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(1):1–15.
- [47] Geri T, Viceconti A, Minacci M, Testa M, Rossetini G. Manual therapy: exploiting the role of human touch. *Musculoskelet Sci Pract* 2019;44:1–20.
- [48] Hróbjartsson A, Forfang E, Haahr MT, Als-Nielsen B, Brorson S. Blinded trials taken to the test: an analysis of randomized clinical trials that report tests for the success of blinding. *Int J Epidemiol* 2007;36:654–63.
- [49] Sackett DL. Commentary: measuring the success of blinding in RCTs: don't, must, can't or needn't? *Int J Epidemiol* 2007;36:664–5.
- [50] 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.
- [51] Hróbjartsson A, Boutron I. Blinding in randomized clinical trials: imposed impartiality. *Clin Pharmacol Ther* 2011;90:732–6.
- [52] Chang K-L, Fillingim R, Hurley RW, Schmidt S. Chronic pain management: nonpharmacological therapies for chronic pain. *FP Essent*; 2015.
- [53] Foster NE, Dziedzic KS, van der Windt DA, Fritz JM, Hay EM. Research priorities for non-pharmacological therapies for common musculoskeletal problems: nationally and internationally agreed recommendations. *BMC Musculoskelet Disord* 2009;10:3.
- [54] Dieppe P. Complex interventions. *Musculoskeletal Care* 2004;2(3):180–6.
- [55] 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(1):1–13.
- [56] Lowe DB, Taylor MJ, Hill SJ. Associations between multimorbidity and additional burden for working-age adults with specific forms of musculoskeletal conditions: a cross-sectional study. *BMC Musculoskelet Disord* 2017;18(1):1–14.
- [57] Duffield SJ, Ellis BM, Goodson N, Walker-Bone K, Conaghan PG, Margham T, et al. The contribution of musculoskeletal disorders in multimorbidity: implications for practice and policy. *Best Pract Res Clin Rheumatol* 2017;31(2):129–44.
- [58] Treweek S, Zwarenstein M. Making trials matter: pragmatic and explanatory trials and the problem of applicability. *Trials* 2009;10:37.
- [59] Negrini S, Arienti C, Pollet J, Schulz KF, Ravaud P. Clinical replicability of rehabilitation interventions in randomized controlled trials reported in main journals is inadequate. *J Clin Epidemiol* 2019;114:108–17.
- [60] Yamato TP, Maher CG, Saragiotto BT, Hoffmann TC, Moseley AM. How completely are physiotherapy interventions described in reports of randomised trials? *Physiotherapy* 2016;102(2):121–6.
- [61] Boutron I, Altman DG, Moher D, et al. 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.
- [62] Groeneweg R, Rubinstein SM, Oostendorp RAB, Ostelo RWJG, Van Tulder MW. Guideline for reporting interventions on spinal manipulative therapy: consensus on interventions reporting criteria list for

- spinal manipulative therapy (CIRCLe SMT). *J Manipulative Physiol Ther* 2017;40(2):61–70.
- [63] Page P, Hoogenboom B, Voight M. Improving the reporting of therapeutic exercise interventions in rehabilitation research. *Int J Sports Phys Ther* 2017;12(2):297–304.
- [64] Hefford C, Abbott JH, Baxter GD, Arnold R. Outcome measurement in clinical practice: practical and theoretical issues for health related quality of life (HRQOL) questionnaires. *Phys Ther Rev* 2013;16(3):155–67.
- [65] Gagnier JJ, Moher D, Boon H, Beyene J, Bombardier C. Investigating clinical heterogeneity in systematic reviews: a methodologic review of guidance in the literature. *BMC Med Res Methodol* 2012;12:1.
- [66] Smith T. The challenges of heterogeneity across physiotherapy trials when considering meta-analysis. *Phys Ther Rev* 2013;19(1):39–40.
- [67] 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(2):119–25.
- [68] Calvert M, Blazeby J, Altman DG, Revicki DA, Moher D, Brundage MD. Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. *JAMA* 2013.

4.2 Resultados del tercer trabajo

4.3.1 Descripción de los artículos

Los 120 ensayos incluyeron un total de 12.954 participantes (mediana = 60, rango de 6 a 1.340) y evaluaron el impacto de seis modalidades de terapias manuales: técnicas de tejidos blandos (29%); técnicas de manipulación espinal (23%); movilización articular (21%); tratamientos quiroprácticos (12%); acupresión-reflexología (9%) y tratamiento de manipulación osteopática (6%). El tipo de ensayos más frecuentes incluidos en el estudio fueron los relacionados con problemas de columna como lumbalgia (21%) y dolor de cuello (16%), seguidos de ensayos realizados en sujetos sanos o asintomáticos (12,5%). De los 120 artículos analizados, el 81% correspondió a estudios unicéntricos y el 19% a estudios multicéntricos. El tratamiento consistió en más de una sesión en el 70% de los estudios. Considerando todas las categorías, la concordancia entre pares de revisores en la etapa piloto osciló entre 0,737 y 0,853, lo que se interpreta como una concordancia 'sustancial' o 'casi perfecta'¹³³.

4.3.2 Análisis del reporte de resultados en el texto

El 40% de los ECA identificó explícitamente la variable de resultado principal y el 47,5% reportó haber realizado un cálculo del tamaño de la muestra. En el 46,7% de los ensayos, el reporte de las comparaciones entre grupos fue completo (incluido el tamaño del efecto y su precisión). Solo el 29,2% utilizó el intervalo de confianza como medida de precisión. El 58% de los ensayos reportó diferencias significativas en los resultados y el 30,8% reportó un valor de relevancia clínica para al menos una variable del estudio. La comparación de los siete elementos clave en el reporte de resultados a lo largo del tiempo, así como la representación gráfica del porcentaje de cumplimiento, queda representado en las tablas y gráficos del artículo.

4.3.3 Análisis del reporte gráfico de resultados

El 47% de los ensayos clínicos reportaron gráficamente los resultados de la variable primaria. Los tipos de gráficos utilizados fueron medidas repetidas (62,5%), histogramas comparativos (26,8%), diagramas de caja (7,1%) y Forest-plot (3,6%). Ninguna de las figuras estaba completa según los criterios previamente establecidos¹³⁴. Solo el 19,6% de los gráficos eran autoexplicativos. El número de sujetos representados en el gráfico faltaba en el 92,9% de los ensayos, el título de los gráficos estaba incompleto en el 83,9% y el 66,1% tenía problemas con la claridad visual.

4.3.4 Tercera publicación

Núñez-Cortés, Rodrigo, Alvarez, Gerard, Pérez-Bracchiglione Javier, Cabanas-Valdés Rosa, Calvo-Sanz Jordi, Bonfill Xavier, Urrútia Gerard. ***Reporting results in manual therapy clinical trials: A need for improvement.*** Int J Osteopath Med. 12 July 2021. doi:<https://doi.org/10.1016/j.ijosm.2021.06.002>. IF (2021): 2.14



Contents lists available at ScienceDirect

International Journal of Osteopathic Medicine

journal homepage: www.elsevier.com/locate/ijosm

Reporting results in manual therapy clinical trials: A need for improvement

Rodrigo Núñez-Cortés^{a,1,*}, Gerard Alvarez^{b,c,1,2,**}, Javier Pérez-Bracchiglione^d,
Rosa Cabanas-Valdés^e, Jordi Calvo-Sanz^{e,f}, Xavier Bonfill^{b,g}, Gerard Urrutia^{b,g}

^a Department of Physical Therapy, Faculty of Medicine, University of Chile, Santiago, Chile^b Iberoamerican Cochrane Centre - Biomedical Research Institute Sant Pau, IIB Sant Pau, Barcelona, Spain^c Centre for Osteopathic Medicine - C.O.M.E. Collaboration, Spain National Center, Barcelona, Spain^d Interdisciplinary Centre for Health Studies (CIESAL), Universidad de Valparaíso, Valparaíso, Chile^e Universitat Internacional de Catalunya, Physiotherapy Department, Faculty of Medicine and Health Sciences, Barcelona, Spain^f Departamento de Rehabilitación. Hospital - Centro de Prevención y Rehabilitación Asepeyo. St. Cugat del Vallés, Barcelona, Spain^g CIBER Epidemiología y Salud Pública (CIBERESP), Spain

ARTICLE INFO

Keywords

CONSORT

Manual therapy

Reporting

P values

Randomized controlled trials

ABSTRACT

Background: The number of randomized clinical trials (RCTs) for manual therapy (MT) has increased exponentially in recent years but the quality of reporting is heterogeneous.

Objective: To assess the quality of the reporting of results in RCTs manual therapy, both in the text and in the graphs.

Study design: Methodological review.

Methods: We reviewed a random sample of 120 RCTs in MT published between 2000 and 2020 in indexed journals. We identified the primary outcome for each trial, and evaluated the completeness and correctness of reporting of results in the text and in the graphs.

Results: Forty per cent of the RCTs explicitly identified the primary outcome and 47.5% reported a sample size calculation. In 46.7% of the trials, the reporting of between groups comparisons was complete (including effect size and precision). Only 29.2% used the confidence interval as a measure of precision. Fifty-eight per cent of the trials reported significant differences in the results, and 30.8% reported a value of clinical relevance for at least one variable of the study. Forty-seven per cent reported the primary outcome graphically but only 19.6% of the graphs were self explanatory and 66.1% had problems of visual clarity.

Conclusions: Our findings suggest that the reporting of the results in MT trials is generally incomplete and graphics are often poor. These shortcomings could affect the interpretation of the results and their application in clinical practice. Improvements are needed in the reporting of results in order to advance clinical practice and research in manual therapy.

* Corresponding author. Department of Physical Therapy, Faculty of Medicine, University of Chile, Avenida Independencia 1027, Independencia, Santiago, Chile.

** Corresponding author. Iberoamerican Cochrane Center C, Sant Antoni Maria Claret 167 Pavelló 18, Planta 0 08025 Barcelona, Catalunya, Spain.

E-mail addresses: r_nunez@uchile.cl (R. Núñez-Cortés), gerardaly@gmail.com (G. Alvarez), javier.perez.b@gmail.com (J. Pérez-Bracchiglione), rosacabanas@uic.es (R. Cabanas-Valdés), jcalvosanz@gmail.com (J. Calvo-Sanz), XBonfill@santpau.cat (X. Bonfill), gurrutia@santpau.cat (G. Urrutia).

¹ RNC and GA contributed equally.

² Gerard Alvarez is a PhD student on Biomedical Research Methodology and Public Health in the Medical Department of the Universitat Autònoma de Barcelona, Barcelona, Spain.

<https://doi.org/10.1016/j.ijosm.2021.06.002>

Received 19 January 2021; Received in revised form 23 June 2021; Accepted 25 June 2021

Available online 12 July 2021

1746-0689/© 2021 Elsevier Ltd. All rights reserved.

Implications for practice

- Our findings suggest that the reporting of results in Manual Therapy trials is generally incomplete and graphics lack completeness and visual clarity.
- Our findings support the need to implement strategies to better comply with reporting guidelines.
- Training in evidenced-based methods with a clear focus on clinical relevance and applicability is needed in the field of Manual Therapy.

Introduction

Manual therapy (MT) is a physical treatment used to treat musculoskeletal pain and disability. It consists of a variety of manual approaches that include joint mobilization, manipulation, massage and soft tissue techniques [1]. The number of randomized controlled trials (RCTs) for MT has increased exponentially in recent years but the quality of methodology and reporting is heterogeneous [2,3]. Several recent systematic reviews [4–12] show that results in MT RCTs are inconclusive, raising the need for studies with better methodology and reporting [13–18]. Comprehensive reporting enables optimal interpretation of the methods and results, facilitating assessment of the implications for clinical practice [19,20]. Moreover, it allows other researchers to learn from prior experiences, overcome previous limitations, and replicate the study to verify the findings when needed or appropriate [21].

A set of guidelines are available to help researchers improve the reporting of the studies. Since 1996, the Consolidated Standards for Reporting Trials (CONSORT statement) have been providing a series of useful resources to improve the quality of reports of clinical trials [22]. However, recent research shows that use of these reporting guidelines has not been optimal [23], leading to poor interpretation and compromising the possibility of using these results in systematic reviews [24, 25]. Moreover, in the field of MT, evidence suggests that the quality of reporting has not improved in recent years [2,26].

Results that provide information about the effects of interventions are a key aspect of RCTs and correct reporting is therefore of utmost importance. The data reported include both the numerical expression of the results in the text (effect and precision measurements) and their representation in graphs or figures. Figures play a key role in communicating findings and facilitating their interpretation. Puhan et al. [27] considered three main requirements for optimal understanding of a graph: first, that data symbols must be clearly distinguishable to allow visual detection, second, that data should be organized in such a way as to facilitate the estimation of the important values and relationships in the data, and third, that there must be a complete explanation of all the elements of the graph and how the data was obtained. However, the presentation or content of many published figures is deficient [28].

Considering that limitations in the reporting of results are associated with biased estimates of treatment effects [29] and given the generally limited quality of reporting in MT RCTs [2,26], the need arises to carry out a complete analysis of the reporting of results and their graphic representation in the MT literature. Therefore, the aim of this study was to assess the quality of the reporting of results in RCTs manual therapy, both in the text and in the graphs. For this purpose, we examined the results sections and specific elements of the methods related to reporting results, in a randomly selected sample of 120 MT studies published between 2000 and 2020.

Methods

Search strategy

We conducted a secondary analysis of a systematic methodological evaluation of a random sample of 100 MT RCTs published between 2000 and 2015 in indexed journals [2]. Criteria for inclusion were that at least one of the interventions (experimental or control) should include some form of MT. Articles not written in English, with designs other than RCTs and studies that referred to posters and oral communications, were excluded. For the purposes of the present study, we extended our sample with 20 additional trials published from 2015 to 2020 to assure the validity of the results and to be representative of articles published up to the present day. A previously validated search filter was used to identify RCTs in the MT field [30]. More details on the search strategy can be found elsewhere [2].

Assessment of reporting

The reporting in the text

A data extraction form agreed upon by four authors (RNC, GA, JPB and GU) was designed. The form was piloted and optimized by four researchers (RNC, JPB, RC and JC) using a random sample of five potentially eligible articles. Regarding the general extraction process, for each article, we determined whether or not the primary outcome was differentiated from the rest of the variables. If the primary outcome was not explicitly declared, the outcome used to calculate the sample size was considered. If neither of these was reported, we selected the first outcome reported in the results section of the study. From this primary outcome, we extracted data regarding: a) the type of outcome (variable); b) the time point of outcome assessment; c) the statistical test used; d) the measure of effect and precision obtained (within and between-group differences); e) the way the result was reported (narratively or graphically); and f) information related to losses to follow-up. We also determined whether the articles explicitly defined the study hypothesis, if they included details of sample size calculation, and if they applied a minimal clinically important difference value (MCID). All data were extracted by three reviewers (RN, RC and JC) and any disagreement was resolved by a fourth reviewer (GA). The evaluation criteria for each item, binary or categorical, are detailed.

In order to assess the evolution of the reporting of the results over time and, due to the lack of a validated statement specifically focused on the reporting of results [31,32], authors agreed on a minimum set of seven categories deemed essential for a complete reporting of RCT results: 1) Primary outcome is explicitly declared (and differentiated from secondary outcomes); 2) Complete information on sample size calculation (including alpha/beta values and standard deviation or minimum detectable change); 3) Complete loss report (including numbers, reasons and imputation criteria of missing values); 4) Intra-group effect measure (experimental group); 5) Intra-group effect measure (control group); 6) Complete comparative measure between groups (effect size and precision); 7) Minimal clinically important difference value.

Graphic report

We identified all RCTs in our sample that presented the primary outcome in a visual format. We identified the RCTs that graphically presented the primary outcome. We determined the type of graph, completeness, and visual clarity according to the recommendations of previous studies [33]. For completeness, we determined whether the graphics included the number of participants, graph and axis titles, axis labels, the meaning of variance, and a self-explanatory presentation. For visual clarity, we examined the graphics for numerical distortion, chart junk, readability, and Others. Further details are provided in [Appendix 1](#). All data were extracted by two reviewers (RNC and JPB) and any disagreement was resolved by a third reviewer (GA).

Statistical analysis

Descriptive statistics were performed using frequencies and percentages for qualitative variables. We calculated agreement between every possible pair of reviewers using Cohen's kappa coefficient and the result was interpreted according to the Landis scale [34]. Chi-square was used to compare the reporting of results by years and one-way ANOVA was used to compare the compliance percentage related to these seven categories. All statistical analyses were performed with SPSS version 22.0 (IBM Corporation, Armonk, NY).

Results

The 120 trials included a total of 12954 participants (median = 60, range 6–1340) and they assessed the impact of six modalities of manual therapies: soft tissue techniques (29%); spinal manipulation techniques (23%); joint mobilization (21%); chiropractic treatments (12%); acupressure-reflexology (9%) and osteopathic manipulative treatment (6%). The most frequent type of trials included in the study were those related to spinal problems such as low back pain (21%) and neck pain (16%), followed by trials conducted in healthy or asymptomatic subjects (12.5%). Of the 120 articles analysed, 81% corresponded to single-centre studies and 19% to multi-centre studies. Treatment consisted of more than one session in 70% of the studies. Considering all categories, the agreement between pairs of reviewers in the pilot stage ranged from

Table 1
Reporting of descriptive information in relation to the study methods.

Hypothesis (n = 120)	
Explicitly stated	35.8% [27.3 to 45.1]
Implicit	40% [31.2 to 49.3]
Not specified	24.4% [16.8 to 32.8]
Primary variable (n = 120)	
Explicitly declared	40% [19.3 to 53.3]
Deductible from sample size	10% [5.2 to 16.8]
Undeclared	50% [40.7 to 59.3]
Variable type (n = 120)	
Continuous	91.7% [85.2 to 95.9]
Dichotomous	5.8% [2.4 to 11.6]
Ordinal	2.5% [0.5 to 7]
Follow-up (n = 120)	
Short term (≤3 months)	70% [61.0 to 78.0]
Long term (>3 months)	30% [22.0 to 39.0]
Statistical analysis (n = 120)	
ANOVA	35.8% [27.3 to 45.1]
Mixed repeated measures model	5.8% [2.4 to 11.6]
Mean difference	1.7% [0.2 to 5.9]
T test	18.3% [11.9 to 26.4]
Chi-Square	5.8% [2.4 to 11.6]
Risk Ratio, Odds Ratio or Hazard Ratio	2.5% [0.5 to 7]
ANCOVA	10.8% [5.9 to 17.8]
Mann – Whitney U test	5% [1.9 to 10.6]
Regression models	6.7% [2.9 to 12.7]
Other	4.2% [1.3 to 9.5]
Undeclared	3.3% [0.9 to 8.3]
Sample size calculation (n = 120)	
Yes	47.5% [38.3 to 56.8]
No	52.5% [43.2 to 61.7]
Specific sample size calculation information (n = 120)	
Alpha value	38.3% [29.6 to 47.6]
Beta value	45.8% [36.7 to 55.2]
Minimal change to be achieved with the intervention	37.5% [28.8 to 46.8]
Standard deviation in continuous variables	12.5% [7.1 to 19.8]
Expected rate or mean in the control group	4.2% [1.3 to 9.5]
Delta margin (for non-inferiority studies)	0%
Imputation criteria (n = 100)	
Yes, defined criteria	24% [16.0 to 33.6]
Not specified	22% [14.3 to 31.4]
Analysis per protocol	23% [15.2 to 32.5]
Not applicable (no losses)	31% [22.1 to 41.0]

Data expressed as a percentage (%) and 95% confidence interval.

Table 2
Reporting specific information regarding results.

Lost to follow up and withdrawals (n = 120)	
Enough information to assess this	83.3% [75.4 to 89.5]
Insufficient information to assess this	16.7% [10.5 to 24.6]
Reasons for losses (n = 100)	
Described	46% [36.0 to 56.3]
Not specified	23% [15.2 to 32.5]
Not applicable (no losses)	31% [22.1 to 41.0]
Measurement of intra-group effect (n = 120)	
Measure of effect (change) is reported	56.7% [47.3 to 65.7]
Only data from each time point is reported	43.3% [34.3 to 52.7]
Comparative measure (n = 120)	
Effect measure and precision	46.7% [37.5 to 56.0]
Precision only (P value)	47.5% [38.3 to 56.8]
Effect measure only	0.8% [0 to 4.6]
No measure of effect nor precision	5% [1.9 to 10.6]
Reported precision measure (n = 120)	
P value	88.3% [81.2 to 93.5]
95% confidence interval	29.2% [21.2 to 38.2]
None	5.8% [2.4 to 11.6]
Statistically significant result (n = 120)	
Yes	58.3% [49.0 to 67.3]
No	41.7% [32.7 to 51.0]
Clinical relevance value (n = 120)	
Yes	30.8% [22.7 to 39.9]
No	58.3% [49.0 to 67.3]
Deductible from sample size	10.8% [5.9 to 17.8]
Clinical relevance result (n = 120)	
Yes	32.5% [24.2 to 41.7]
No or not specified	67.5% [58.3 to 75.8]

Data expressed as a percentage (%) and 95% confidence interval.

0.737 to 0.853, which is interpreted as substantial to almost perfect agreement.

Analysis of the reporting of results in the text and tables

Table 1 provides a detailed description of the study methods and Table 2 shows specific information regarding the reporting of results. Forty per cent of the RCTs explicitly identified the primary outcome and 47.5% reported a sample size calculation. In 46.7% of the trials, the reporting of between groups comparisons was complete (including effect size and precision). Only 29.2% used the confidence interval as a measure of precision. Fifty-eight per cent of the trials reported significant differences in the results, and 30.8% reported a clinical relevance value for at least one variable of the study. Table 3 shows a comparison of seven key elements of the reporting results over time. The compliance percentage related to these seven categories is represented graphically in Fig. 1 (see Table 4).

Analysis of the graphic report of the results

Forty-seven per cent of the clinical trials reported the results for the primary outcome graphically. The types of graphs used were repeated measures (62.5%), comparative histograms (26.8%), box-plots (7.1%) and forest plots (3.6%). None of the figures were complete according to the previously established criteria [33]. Only 19.6% of the graphs were self explanatory. The number of subjects represented in the graph was missing in 92.9% of trials, the title of the graphs was incomplete in 83.9%, and 66.1% had problems with visual clarity. Details of the graphic report are shown in Table 3.

Discussion

To the best of our knowledge, this is the first study to provide a picture of the reporting quality of MT results over a considerable period of time. Our results suggest that the reporting of results is generally incomplete, despite some positive trends identified in the recent

Table 3

Comparison of the reporting of results by years.

Categories	2000–2004 (n = 24)	2005–2009 (n = 31)	2010–2014 (n = 45)	2015–2019 (n = 20)	p value
Primary variable is explicitly declared	11 (45.8%)	12 (38.7%)	19 (42.2%)	6 (30%)	0.732
Sample size calculation (complete information) ^a	7 (29.2%)	10 (32.3%)	15 (33.3%)	19 (50%)	0.476
Complete loss report ^b	11 (45.8%)	15 (48.4%)	27 (60%)	13 (65%)	0.453
Intra-group effect measure (experimental group)	13 (54.2%)	21 (67.7%)	21 (46.7%)	13 (65%)	0.261
Intra-group effect measure (control group)	12 (50%)	20 (64.5%)	17 (37.8%)	12 (60%)	0.108
Complete comparative measure between groups ^c	6 (25%)	7 (22.6%)	13 (28.9%)	9 (45%)	0.350
Clinical relevance threshold	7 (29.2%)	8 (25.8%)	19 (42.2%)	3 (15%)	0.139

Data are expressed in percentages (%).

^a including alpha/beta values and standard deviation or minimum detectable change).^b including numbers, reasons and imputation criteria for missing data.^c including effect size and precision measure.**Table 4**

Evaluation of the graphic report (n = 56).

Completeness	
Number of subjects is discernible for each graph element	7.1 [1.9 to 17.3]
Title	16.1 [7.6 to 28.3]
x axis, y axis titles	80.4 [67.6 to 89.8]
x axis, y axis labels	82.1 [69.6 to 91.1]
Variance meaning defined	35.7 [23.4 to 49.6]
Self explanatory	19.6 [10.2 to 32.4]
Visual clarity (Absence of the following)	
Numerical distortion	87.5 [75.9 to 94.8]
Chart junk	73.2 [59.7 to 84.2]
Readability issues	53.6 [39.7 to 67.0]
Others	89.3 [78.1 to 96.0]

Data expressed as percentages (%) and 95% confidence interval. Evaluation criteria: Number of subjects is discernible for each graph element; Title (explicitly detailed); x axis, y axis titles (explicitly detailed); x axis, y axis labels (i.e. for tick marks or categories, label and units clear); Variance meaning defined (e.g., standard error or 95% CI); Self-explanatory (all defined data elements, including legend); Numerical distortion (scale problems or improperly scaled axes or improperly ranged axes); Chart junk (cross-hatching patterns or dark/thick/unnecessary grid lines or text labels in non-horizontal orientation); Readability issues (error bars too cluttered or superimposition of data elements or display too small to see symbols or numerical/textual redundancy); Others (improperly connected points, labels too small to read or unclear to which graph item label refers).

literature.

These findings, specifically focused on results-related items on MT trials, are in accordance with similar studies in other healthcare disciplines [35–42] and add to the necessity for a more clear guidance on how to report the results of an RCT [31,43,44]. Fortunately, these shortcomings have fostered promising initiatives like the current development of the Instrument for reporting Planned Endpoints in Clinical Trials (InsPECT) [32]. Once available, the InsPECT instrument could be used by individuals responsible for the design, implementation, and reporting of clinical trials and also help scientific journal editors to improve the peer review process of RCTs (<https://www.inspect-statement.org>).

Despite our findings, we want to highlight some positive trends registered in the most recent literature. This includes a more complete reporting on comparative measures between groups, a more complete loss report and sample size calculation. Notwithstanding, in the following discussion, we focus on the incomplete and missing items in order to highlight their relevance and offer some guidance and recommendations for improving future research.

Primary and secondary outcomes

In our sample, fewer than 50% of the articles explicitly stated the primary outcome of the study. CONSORT guidelines have long

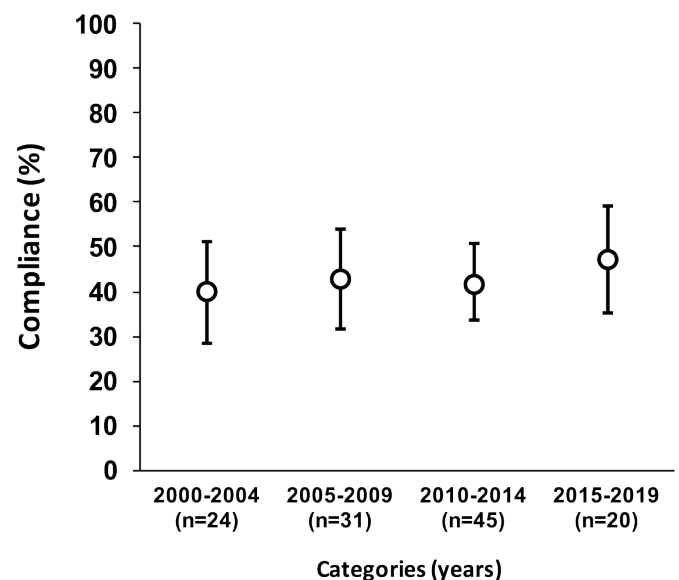
Comparison of the Reporting of Results by Years

Fig. 1. Compliance (%) was defined by seven categories: a) Primary outcome is explicitly declared; b) Complete information on sample size calculation; c) Complete loss report; d) Intra-group effect measure (experimental group); e) Intra-group effect measure (control group); f) Complete comparative measure between groups; g) Minimal clinically important difference value. Data are expressed as mean and 95% confidence interval.

recommended that the primary outcome should be specifically reported and differentiated from secondary outcomes [22]. The primary outcome measure is set at the moment the study is designed and the protocol is drafted. Outcomes are based directly on the primary aim of the study [45] and sample size is calculated from the primary outcome. Clearly stating the primary outcome prevents investigators from cherry-picking significant results and presenting these as the main findings of the study. To avoid this reporting bias, it is strongly recommended to prospectively register the study protocol on a public trial registry.

Sample size, statistical significance and clinical relevance

The under-reported description of how sample size was determined in our study (35.8%) appears to be a common finding in trials in Physical Medicine and Rehabilitation [46,47], limiting readers' ability to properly interpret the results. The literature shows that MT trials are typically based on small sample size [48] and trials that perform an a priori sample size calculation have considerably larger median sample sizes than those that do not [49]. In addition, complete reporting on how sample size was calculated prevents the exclusive reliance on statistical power and also addresses key elements such as the importance of the

selected outcome and the MCID to be detected [8,50,51].

In our sample, only 30.8% of the trials specified MCID either explicitly or within the sample size calculation, and within the 58.3% of trials reporting statistically significant results, only 35.7% were clinically relevant. Due to the nature of MT interventions, clinical trials often aim to show effectiveness by applying a pragmatic approach to the study design. In this scenario, the reporting of a MCID provides a valuable reference to evaluate the potential impact of the findings on clinical practice [52].

Effect estimates and precision

Complete reporting of effect size and its precision both within and between-groups is important to correctly interpret the results and to provide a broader understanding of the effect of an intervention [43]. However, while all trials in our sample reported intra-group effects in some way, only 46.7% accurately reported effect size and its precision in between-group differences. This lack of information raises important clinical considerations in fields such as MT where nonspecific effects play an important role in the overall effect [53–55]. In this regard, in the JOSPT series “Linking evidence with practice” Steve Kamper recently stated: “The former [within-group] includes natural recovery, regression to the mean, nonspecific effects, and treatment effects. The latter [between-group] is the treatment effect” [56]. Therefore, especially in explanatory trials, the inclusion of both estimates must be seen as good reporting practice.

Another important aspect relates to the completeness of reporting the effects of outcomes. Forty-eight per cent only reported the p-value for the between-group comparison. The p-value was used in 88.3% of the cases but the confidence interval was reported in only 29.2% (Table 2). However, a positive trend towards an improvement has been noticed in most recent MT trials (Table 3). Although p values are easily calculated, they often lead to misunderstandings [57–59]. Furthermore, they are also often used incorrectly [60] and may overestimate treatment effects [61]. The reporting of p values alone does not therefore provide a complete interpretation of results [62]. Using confidence intervals as a precision estimate may lead to a better clinical interpretation of the literature in rehabilitation and could also help to establish confidence in the real magnitude of the observed effect [50,63]. In order to improve statistical interpretation and reporting, Greenland et al. published a set of useful guidelines [59].

Loss to follow-up and missing data

Complete assessment of all participants in an RCT is crucial for a sound interpretation of results. However, a number of participants will be lost to follow up or not analysed for some reason at the end of the study, potentially leading to attrition bias. Furthermore, data that are missing can affect the power of the study [64]. In 83.3% of the articles in our sample, enough information was reported to assess whether a loss to follow-up occurred. The number of randomly allocated participants and information on missing data, dropouts, and losses to follow-up are generally well reported in MT trials [2] and our results show again a positive trend in most recent RCTs. However, on further analysing this information, we found that the reasons for loss to follow-up were not reported in 23% of the cases and the imputation criteria of missing data in the analysis were not described in 22% of the cases. Assessment of clinical outcomes and treatment efficacy depends on reliable follow-up information [65]. As recommended in the CONSORT guidelines and extensions, authors should include a patient flow diagram with the number of patients lost to follow-up, the reasons for loss to follow-up, and the number of patients analysed [66,67]. In this scenario, the authors should clearly state the method for handling missing data and the statistical model used to analyse the trial. The review published by Armijo-Olivo et al. offers a very useful guide on how to deal with missing data in clinical research [64].

Graphical report

One of the main shortcomings in RCTs in general is the graphic presentation of results [68,69]. Figures are commonly used tools to present data for scientific communication [70]. The visual impact helps to transmit findings to the reader [71], aiding in their interpretation and making a lasting impression of the results [70]. Pocock et al. [28] observed that the main types of figures used in clinical trials are: flow diagrams, Kaplan Meier diagrams, Forest diagrams, and repeated measurements over time. It is not surprising that the most common type of graph in our sample was the repeated measure graph. This is because continuous variables are those most frequently used in MT studies.

The majority of graphs published in top medical journals fail to display complete data [33]. Furthermore, areas of low performance are similar regardless of the study field. Our results represent a first approach to evaluating these characteristics in the field of MT and they show the same shortcomings. In conclusion, our findings support the need to implement strategies that have proven to be effective in the use of reporting guidelines [72–74]. At the same time, training in evidence-based methods should be promoted among researchers in the MT field, focusing clearly on clinical relevance and applicability.

Strengths and limitations

The main strength of this review relates to the large data set used in the analysis. However, the lack of formal validity tests for our sampling method could have introduced potential biases, especially concerning changes over time. Furthermore, the random sampling method used might have led to some particular MT disciplines being less represented than others. However, our results should be interpreted as a general picture of the quality of MT results reporting over a considerable period of time instead of a particular discipline-based analysis.

Another potential limitation relates to the tool used for data extraction. As a non-validated tool, some inconsistencies could have occurred during the data extraction. Notwithstanding, several actions were implemented to reduce this possibility. First, the data extraction form was designed through agreement between all the research team that includes experts both in the field of research methodology and in MT. Second, we piloted the form on a sample of 5 RCTs, reaching a high agreement between the reviewers. In any case, the aim of the study was not to develop an instrument to evaluate the reporting of results but rather to draw attention to the relevance of complete and sound reporting. Finally, the subanalysis made to explore potential differences over time should be considered cautiously. This assessment was based on the reporting of seven items deemed to be essential by the research team. Although the majority of these items were extracted from reporting guidelines, some of them were selected *ad-hoc* according to the authors judgement.

Conclusions

Our findings suggest that the reporting of results in studies of MT is generally incomplete, limiting comprehensive interpretation. The main deficiencies are related to the identification of the primary outcome, the reporting of sample size calculation, effect estimates and precision, and missing data. Our results also show that the graphic presentation of results often lacks completeness and visual clarity. Improvements in the reporting of results are necessary to generate advances in the practice and research of MT interventions.

Author contributions

RNC, GA, JPB and GU were all involved in the conception and design of the study. RNC, RC, JC and GA were responsible for recruitment and data collection. RNC, GA, JPB, RC, JC, XB and GU contributed to the writing and revision of the manuscript and approved the final version of

the manuscript. GU is the study guarantor.

Ethical approval

Not applicable.

Funding

This research did not receive any specific grant from funding

agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

All authors declare that there are no financial or personal conflicts of interest regarding the publication of this paper.

Appendix 1. Data extraction form

Is the study hypothesis specified?
Yes, explicitly
No, but it is implicit
Not specified (cannot be determined)

If the hypothesis can be determined, it is:
Superiority
Non-inferiority

Is a primary variable described?
Yes (explicitly specified)
No (but it can be identified from the sample size calculation)
No

Variable type (n = 100)
Continuous
Dichotomous
Ordinal category

Follow-up (n = 100)
Short term (≤ 3 months)
Long term (> 3 months)

Statistical analysis (n = 100)
ANOVA
Mixed repeated measures model
Mean difference
T test
Chi-Square
Risk Ratio, Odds Ratio or Hazard Ratio
ANCOVA
Mann-Whitney U test
Regression models
Other

Was the result for the main variable presented graphically?
Yes
No

Was a sample size calculation performed?
Yes
No

What information is given to evaluate how this calculation was made?
Alpha value
Beta value
Minimal change to be achieved with the intervention
Standard deviation in continuous variables
Expected rate or mean in the control group
Delta margin (for non-inferiority studies)

Is there sufficient information to determine whether there was a loss of subjects during the study?
Yes
No

In the event that some subjects were lost to follow-up, did they appear quantified (n°) in the article?
Yes
No
Not applicable (no losses)

If some subjects were lost to follow-up, were reasons described?
Yes
No
Not applicable (no losses)

If some subjects were lost to follow up, were the criteria for imputation of these missing data described?
Yes, analysis by protocol
Yes, other reasons
No
Not applicable (no losses)

Is the measure of the intra-group effect reported?
Measure of effect is reported
Only data from each evaluation is reported

(continued on next page)

(continued)

Is a Comparative Effect Measurement reported?

No effect or precision measure (p-value)
 No measure of effect, but p-value
 Effect measure without p-value
 Measure of effect and p-value

According to the authors, were the results favourable? (significant difference)

Yes
 No

Regarding the results, is a value of clinical relevance specified?

Yes, explicitly
 Yes, a magnitude of change is mentioned in the sample calculation
 No

Based on the previous question: was the result clinically relevant?

Yes
 No
 A threshold of clinical relevance is not specified

Regarding completeness of graphics. Are the following characteristics met? (Yes or no)

Number of subjects is discernible for each graph element
 Title (explicitly detailed) x axis, y axis titles (explicitly detailed)
 x axis, y axis labels (i.e., for tick marks or categories, label and units clear)
 Variance meaning defined (e.g., standard error or 95% CI)
 Self-explanatory (all defined data elements, including legend)

Regarding the visual clarity of graphics, are the following characteristics absent (Yes or no)

Numerical distortion (scale problems or improperly scaled axes or improperly ranged axes)
 Chart junk (cross-hatching patterns or dark/thick/unnecessary grid lines or text labels in nonhorizontal orientation)
 Readability issues (error bars too cluttered or superimposition of data elements or display too small to see symbols or numeric/textual redundancy)
 Others (improperly connected points, labels too small to read or unclear to which graph item label refers).

References

- [1] French HP, Brennan A, White B, Cusack T. Manual therapy for osteoarthritis of the hip or knee – a systematic review. *Man Ther* 2011;16:109–17.
- [2] 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>.
- [3] Shiwa SR, Moseley AM, Maher CG, Costa LOP. Language of publication has a small influence on the quality of reports of controlled trials of physiotherapy interventions. *J Clin Epidemiol* 2013;66:78–84. <https://doi.org/10.1016/j.jclinepi.2012.08.004>.
- [4] Zavala-González J, Pavez-Baeza F, Gutiérrez-Espinoza H, Olguín-Huerta C. The effectiveness of joint mobilization techniques for range of motion in adult patients with primary adhesive capsulitis of the shoulder: a systematic review and meta-analysis. *Medwave* 2018;18:e7265.
- [5] Minkalis AL, Vining RD, Long CR, Hawk C, de Luca K. A systematic review of thrust manipulation combined with one conservative intervention for rotator cuff and related non surgical shoulder conditions. *J Can Chiropr Assoc* 2018;62:5–17.
- [6] Carnes D, Plunkett A, Ellwood J, Miles C. Manual therapy for unsettled, distressed and excessively crying infants: a systematic review and meta analyses. *BMJ Open* 2018;8:e019040.
- [7] de Luca KE, Fang SH, Ong J, Shin K-S, Woods S, Tuchin PJ. The effectiveness and safety of manual therapy on pain and disability in older persons with chronic low back pain: a systematic review. *J Manip Physiol Ther* 2017;40:527–34.
- [8] Bizzarri P, Buzzatti L, Cattrysse E, Scafoglieri A. Thoracic manual therapy is not more effective than placebo thoracic manual therapy in patients with shoulder dysfunctions: a systematic review with meta-analysis. *Musculoskeletal Sci Pract* 2018;33:1–10.
- [9] Laimi K, Mäkilä A, Bärlund E, Katajapuu M, Oksanen A, Seikkula V, et al. Effectiveness of myofascial release in treatment of chronic musculoskeletal pain: a systematic review. *Clin Rehabil* 2018;32:440–50.
- [10] Wu J, Wang Y, Wang Z. The effectiveness of massage and touch on behavioural and psychological symptoms of dementia: a quantitative systematic review and meta-analysis. *J Adv Nurs* 2017;73:2283–95.
- [11] Smith CA, Levett KM, Collins CT, Dahlen HG, Ee CC, Suganuma M. Massage, reflexology and other manual methods for pain management in labour. *Cochrane Database Syst Rev* 2018;3:CD009290.
- [12] Furlan AD, Giraldo M, Baskwill A, Irvin E, Imamura M. Massage for low-back pain. *Cochrane Database Syst Rev* 2015. <https://doi.org/10.1002/14651858.cd001929.pub3>.
- [13] Westad K, Tjøstølvsen F, Hebron C. The effectiveness of Mulligan's mobilisation with movement (MWM) on peripheral joints in musculoskeletal (MSK) conditions: a systematic review. *Musculoskeletal Sci Pract* 2018. <https://doi.org/10.1016/j.msksp.2018.12.001>.
- [14] Lee N-W, Kim G-H, Heo I, Kim K-W, Ha I-H, Lee J-H, et al. Chuna (or tuina) manual therapy for musculoskeletal disorders: a systematic review and meta-analysis of randomized controlled trials. *Evid Based Compl Alternat Med* 2017;8218139. 2017.
- [15] Xu Q, Chen B, Wang Y, Wang X, Han D, Ding D, et al. The effectiveness of manual therapy for relieving pain, stiffness, and dysfunction in knee osteoarthritis: a systematic review and meta-analysis. *Pain Physician* 2017;20:229–43.
- [16] Mischke JJ, Jayaseelan DJ, Sault JD, Emerson Kavchak AJ. The symptomatic and functional effects of manual physical therapy on plantar heel pain: a systematic review. *J Man Manip Ther* 2017;25:3–10.
- [17] Krul M, van der Wouden JC, Kruithof EJ, van Suijlekom-Smit LW, Koes BW. Manipulative interventions for reducing pulled elbow in young children. *Cochrane Database Syst Rev* 2017;7:CD007759.
- [18] Gross A, Langevin P, Burnie SJ, Bédard-Brochu M-S, Empey B, Dugas E, et al. Manipulation and mobilisation for neck pain contrasted against an inactive control or another active treatment. *Cochrane Database Syst Rev* 2015. <https://doi.org/10.1002/14651858.cd004249.pub4>.
- [19] Silva TM, Costa LCM, Costa LOP. Evidence Based Practice: a survey regarding behavior, knowledge, skills, resources, opinions and perceived barriers of Brazilian physical therapists from São Paulo state. *Braz J Phys Ther* 2015;19:294–303.
- [20] da Silva TM, Costa L da CM, Garcia AN, Costa LOP. What do physical therapists think about evidence-based practice? A systematic review. *Man Ther* 2015;20:388–401.
- [21] Straulino S. Reconstruction of Galileo Galilei's experiment: the inclined plane. *Phys Educ* 2008;43:316–21. <https://doi.org/10.1088/0031-9120/43/3/012>.
- [22] Altman DG. Better reporting of randomised controlled trials: the CONSORT statement. *BMJ* 1996;313:570–1.
- [23] Cauley L, Catalá López F, Whelan J, Khoury M, Ferraro J, Cheng W, et al. Reporting guidelines of health research studies are frequently used inappropriately. *J Clin Epidemiol* 2020;122:87–94.
- [24] 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.
- [25] Moseley AM, Herbert RD, Maher CG, Sherrington C, Elkins MR. Reported quality of randomized controlled trials of physiotherapy interventions has improved over time. *J Clin Epidemiol* 2011;64:594–601.
- [26] 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.
- [27] Puhan MA, ter Riet G, Eichler K, Steurer J, Bachmann LM. More medical journals should inform their contributors about three key principles of graph construction. *J Clin Epidemiol* 2006;59:1017–22.
- [28] Pocock SJ, Trivison TG, Wruck LM. Figures in clinical trial reports: current practice & scope for improvement. *Trials* 2007;8:36.
- [29] Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001;134:663–94.
- [30] Pillastrini P, Vanti C, Curti S, Mattioli S, Ferrari S, Violante FS, et al. Using PubMed search strings for efficient retrieval of manual therapy research literature. *J Manip Physiol Ther* 2015;38:159–66.
- [31] Butcher NJ, Monsour A, Mew EJ, Szatmari P, Pierro A, Kelly LE, et al. Improving outcome reporting in clinical trial reports and protocols: study protocol for the Instrument for reporting Planned Endpoints in Clinical Trials (InsPECT). *Trials* 2019;20:161.

- [32] Butcher NJ, Mew EJ, Saeed L, Monsour A, Chee A Tow A, Chan A W, et al. Guidance for reporting outcomes in clinical trials: scoping review protocol. *BMJ Open* 2019;9:e023001.
- [33] Chen JC, Cooper RJ, McMullen ME, Schriger DL. Graph quality in top medical journals. *Ann Emerg Med* 2017;69:453. 61.e5.
- [34] Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
- [35] Bhaloo Z, Adams D, Liu Y, Hansraj N, Hartling L, Terwee CB, et al. Primary Outcomes Reporting in Trials (PORTal): a systematic review of inadequate reporting in pediatric randomized controlled trials. *J Clin Epidemiol* 2017;81: 33–41.
- [36] Sinha IP, Altman DG, Beresford MW, Boers M, Clarke M, Craig J, et al. Standard 5: selection, measurement, and reporting of outcomes in clinical trials in children. *Pediatrics* 2012;129(Suppl 3):S146–52.
- [37] Mayo-Wilson E, Fusco N, Li T, Hong H, Canner JK, Dickersin K, et al. Multiple outcomes and analyses in clinical trials create challenges for interpretation and research synthesis. *J Clin Epidemiol* 2017;86:39–50.
- [38] Mantziari S, Demartines N. Poor outcome reporting in medical research; building practice on spoilt grounds. *Ann Transl Med* 2017;5:S15.
- [39] Dwan K, Gamble C, Williamson PR, Kirkham JJ. Reporting Bias Group. Systematic review of the empirical evidence of study publication bias and outcome reporting bias - an updated review. *PLoS One* 2013;8:e66844.
- [40] Hall NJ, Kapadia MZ, Eaton S, Chan WWY, Nickel C, Pierro A, et al. Outcome reporting in randomised controlled trials and meta-analyses of appendicitis treatments in children: a systematic review. *Trials* 2015;16:275.
- [41] Saldanha IJ, Dickersin K, Wang X, Li T. Outcomes in Cochrane systematic reviews addressing four common eye conditions: an evaluation of completeness and comparability. *PLoS One* 2014;9:e109400.
- [42] Chan A-W, Altman DG. Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors. *BMJ* 2005;330:753.
- [43] Froud R, Underwood M, Eldridge S. Improving the reporting and interpretation of clinical trial outcomes. *Br J Gen Pract* 2012;62:e729–31. <https://doi.org/10.3399/bjgp12x657008>.
- [44] Heneghan C, Goldacre B, Mahtani KR. Why clinical trial outcomes fail to translate into benefits for patients. *Trials* 2017;18:122.
- [45] Vetter TR, Mascha EJ. Defining the primary outcomes and justifying secondary outcomes of a study: usually, the fewer, the better. *Anesth Analg* 2017;125:678–81.
- [46] Abdul Latif I, Daud Amadera JE, Pimentel D, Pimentel T, Fregni F. Sample size calculation in physical medicine and rehabilitation: a systematic review of reporting, characteristics, and results in randomized controlled trials. *Arch Phys Med Rehabil* 2011;92:306–15.
- [47] Castellini G, Gianola S, Bonovas S, Moja L. Improving power and sample size calculation in rehabilitation trial reports: a methodological assessment. *Arch Phys Med Rehabil* 2016;97:1195–201.
- [48] Koes BW. How to evaluate manual therapy: value and pitfalls of randomized clinical trials. *Man Ther* 2004;9:183–4.
- [49] Rubinstein SM, van Eekelen R, Oosterhuis T, de Boer MR, Ostelo RWJG, van Tulder MW. The risk of bias and sample size of trials of spinal manipulative therapy for low back and neck pain: analysis and recommendations. *J Manip Physiol Ther* 2014;37:523–41.
- [50] Page P. Beyond statistical significance: clinical interpretation of rehabilitation research literature. *Int J Sports Phys Ther* 2014;9:726–36.
- [51] Armijo-Olivo S. The importance of determining the clinical significance of research results in physical therapy clinical research. *Braz J Phys Ther* 2018;22:175–6.
- [52] Armijo-Olivo S, Warren S, Fuentes J, Magee DJ. Clinical relevance vs. statistical significance: using neck outcomes in patients with temporomandibular disorders as an example. *Man Ther* 2011;16:563–72.
- [53] Rossetini G, Carlino E, Testa M. Clinical relevance of contextual factors as triggers of placebo and nocebo effects in musculoskeletal pain. *BMC Musculoskel Disord* 2018;19. <https://doi.org/10.1186/s12891-018-1943-8>.
- [54] 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.
- [55] 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.
- [56] Kamper SJ. Interpreting outcomes 1—change and difference: linking evidence to practice. *J Orthop Sports Phys Ther* 2019;49:357–8.
- [57] Goodman S. A dirty dozen: twelve P-value misconceptions. *Semin Hematol* 2008; 45:135–40. <https://doi.org/10.1053/j.seminhematol.2008.04.003>.
- [58] Oliveri RS, Gluud C, Wille-Jørgensen PA. Hospital doctors' self-rated skills in and use of evidence-based medicine - a questionnaire survey. *J Eval Clin Pract* 2004;10: 219–26.
- [59] Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, et al. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *Eur J Epidemiol* 2016;31:337–50.
- [60] Bassler D, Briel M, Montori VM, Lane M, Glasziou P, Zhou Q, et al. Stopping randomized trials early for benefit and estimation of treatment effects: systematic review and meta-regression analysis. *J Am Med Assoc* 2010;303:1180–7.
- [61] Bleakley C, Reijgers J, Smoliga JM. Many high quality randomized controlled trials in sports physical therapy are making false positive claims of treatment effect: a systematic survey. *J Orthop Sports Phys Ther* 2020;50:104–9. <https://doi.org/10.2519/jospt.2020.9264>.
- [62] Fethney J. Statistical and clinical significance, and how to use confidence intervals to help interpret both. *Aust Crit Care* 2010;23:93–7.
- [63] Kamper SJ. Confidence intervals: linking evidence to practice. *J Orthop Sports Phys Ther* 2019;49:763–4.
- [64] Armijo-Olivo S, Warren S, Magee D. Intention to treat analysis, compliance, drop-outs and how to deal with missing data in clinical research: a review. *Phys Ther Rev* 2009;14:36–49.
- [65] von Allmen RS, Weiss S, Tevaearai HT, Kuemmerli C, Tinner C, Carrel TP, et al. Completeness of follow-up determines validity of study findings: results of a prospective repeated measures cohort study. *PLoS One* 2015;10:e0140817.
- [66] 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. 10 1136/bmj.c869.
- [67] Akl EA, Briel M, You JJ, Sun X, Johnston BC, Busse JW, et al. Potential impact on estimated treatment effects of information lost to follow-up in randomised controlled trials (LOST IT): systematic review. *BMJ* 2012;344:e2809.
- [68] Schriger DL, Sinha R, Schroter S, Liu PY, Altman DG. From submission to publication: a retrospective review of the tables and figures in a cohort of randomized controlled trials submitted to the British Medical Journal. *Ann Emerg Med* 2006;48. 750–6, 756.e1–21.
- [69] Schriger DL, Cooper RJ. Achieving graphical excellence: suggestions and methods for creating high-quality visual displays of experimental data. *Ann Emerg Med* 2001;37:75–87.
- [70] Pocock SJ, Trivison TG, Wruck LM. How to interpret figures in reports of clinical trials. *BMJ* 2008;336:1166–9.
- [71] Cleveland WS. The elements of graphing data. 1992.
- [72] Blanco D, Altman D, Moher D, Boutron I, Kirkham JJ, Cobo E. Scoping review on interventions to improve adherence to reporting guidelines in health research. *BMJ Open* 2019;9:e026589.
- [73] Hopewell S, Boutron I, Altman DG, Barbour G, Moher D, Montori V, et al. Impact of a web based tool (WebCONSORT) to improve the reporting of randomised trials: results of a randomised controlled trial. *BMC Med* 2016;14:199.
- [74] Barnes C, Boutron I, Giraudeau B, Porcher R, Altman DG, Ravaud P. Impact of an online writing aid tool for writing a randomized trial report: the COBWEB (Consort-based WEB tool) randomized controlled trial. *BMC Med* 2015;13:221.

5. DISCUSIÓN

5. Discusión

5.1 Principales resultados derivados de los trabajos

El primer y tercer trabajo que conforma esta tesis doctoral han permitido obtener datos objetivos relacionados con la calidad del reporte de la literatura científica en el campo de la TM. El análisis se ha dirigido de forma específica a aquellos elementos metodológicamente determinantes en cuanto a la validez, la fiabilidad de los estudios, la descripción de las intervenciones y el reporte de resultados. Es por ello que este análisis del reporte constituye en si mismo una valoración indirecta sobre la calidad de la evidencia en este campo. En el segundo trabajo, se han identificado las principales limitaciones (metodológicas y no-metodológicas) que afectan a la literatura científica en TM. Conjuntamente, se ha obtenido un conocimiento exhaustivo sobre las principales debilidades que afectan a este cuerpo de evidencia y cuáles son las dificultades (metodológicas y contextuales) que se presentan en el momento de realizar investigación experimental en TM.

Los resultados del primer trabajo¹³⁵ indican que la calidad general del reporte de ECAs en TM es deficiente y que este no ha mejorado con el tiempo, salvo en unos pocos aspectos. Esta deficiente calidad es especialmente preocupante en el caso del reporte de los resultados de los estudios tal y como se evidencia en el tercer trabajo de este compendio¹³⁶. El reporte de resultados en ECAs de TM es generalmente incompleto, lo que limita de forma importante la correcta interpretación de los mismos y, en consecuencia, la utilidad e impacto de la investigación en la práctica clínica. Se evidencia también una deficiente presentación gráfica de estos resultados que, a menudo, es incompleta y carece de claridad visual. En su conjunto, estos hallazgos sugieren que las guías de reporte (especialmente CONSORTnpt) no han tenido un impacto significativo en la calidad de los informes años después de su publicación y adopción por parte de numerosas revistas científicas. Este dato, junto con otras limitaciones reportadas sobre el uso y la aplicabilidad de estas guías, nos hacen sugerir diferentes estrategias para mejorar su usabilidad y cumplimiento. Entre ellas, destacamos y proponemos la creación de guías específicas para intervenciones no farmacológicas concretas (como podría ser la TM) que permitan un reporte más adaptado a cada campo específico dentro de las llamadas intervenciones no-farmacológicas.

En relación al segundo trabajo¹³⁷, nuestros resultados indican que al menos dos limitaciones diferentes se reconocen sistemáticamente en los ECAs de TM. Además, el análisis de los argumentos reportados por los autores ayuda a comprender las dificultades específicas que enfrentan los investigadores en el ámbito de las TMs y, a su vez, permite establecer estrategias, recomendaciones o pautas para solucionarlas. Aunque reportadas en una proporción diferente, pudimos diferenciar LARs con mayor impacto en la validez interna del ensayo de otras no directamente relacionadas con el diseño del estudio o que reflejan aspectos más relacionados con su realización. De acuerdo con su naturaleza, las limitaciones pueden tener un impacto en la confianza que podamos tener en las estimaciones del efecto de las intervenciones evaluadas en los estudios o en sus conclusiones, y esto podría tener algunas implicaciones en la práctica. Los investigadores deben asegurar el reconocimiento de aquellas limitaciones que puedan implicar amenazas a la validez del estudio al preparar sus manuscritos para informar sobre los resultados de un ECA. Por otro lado, con relación a las guías de reporte, se aconseja incluir directrices más específicas para incrementar la precisión en el reconocimiento de las limitaciones, diferenciando cuáles deben

ser declaradas de forma obligatoria de aquellas otras que son altamente recomendables. Las revistas también deberían facilitar herramientas para evaluar las limitaciones y el potencial impacto que estas pueden tener. Todo ello, debería ser contemplado de una forma más detallada durante el proceso de revisión por pares.

5.2 Discusión de los elementos más relevantes

El análisis exhaustivo del reporte y de las LARs de una muestra amplia y representativa de ECAs en el campo de la TM ha permitido establecer un listado de elementos metodológicamente “conflictivos” en este campo. Haber conseguido esta información es fundamental para poder avanzar en el propósito de desarrollar estrategias metodológicas que puedan incrementar la calidad de los estudios experimentales en TM. A continuación, se presenta la discusión detallada de estos elementos, de sus razones así como las posibles soluciones metodológicas y recomendaciones para futuras investigaciones. Toda esta discusión se presenta de forma esquematizada al final de esta sección en la **Tabla 3**.

5.2.1 Tamaño de la muestra

Los ECA en TM con frecuencia se realizan con tamaños de muestra pequeños³¹. La investigación en TM a menudo se lleva a cabo en centros privados y servicios ambulatorios, lo que limita las posibilidades de reclutamiento dando lugar a estudios con muestras reducidas de participantes. Por ejemplo, en nuestra muestra de artículos, sólo el 31% de los ECA tenían muestras de más de 100 participantes. En consecuencia, no es de extrañar que la LAR mayormente identificada en el segundo trabajo de esta tesis estuviera relacionada con el tamaño de la muestra (47,5%). Dentro de esta categoría, la principal razón dada por los autores fue que los estudios carecían de poder estadístico para detectar diferencias de efecto entre los grupos de estudio (87,7%)¹³⁷. Por otro lado, es importante considerar que muchas intervenciones no-farmacológicas (y en especial la TM) se utilizan habitualmente para tratar enfermedades crónicas¹³⁸⁻¹⁴⁰. El tratamiento puede durar varias semanas, por lo que la evaluación de los resultados por necesidad incluye un seguimiento a largo plazo. En este escenario, los investigadores deben ser conscientes del hecho de que una muestra pequeña o insuficiente en un ECA con seguimiento a largo plazo puede verse potencialmente afectada por el sesgo de desgaste.

La literatura científica evidencia que el tamaño de la muestra está asociado con su cálculo previo, de tal manera que los ensayos que realizan un cálculo formal del tamaño de la muestra *a priori* tienen de media tamaños de muestra considerablemente mayores que los que no lo hacen¹⁴¹. Sin embargo, a pesar de la mejora observada a lo largo del tiempo, el reporte del cálculo del tamaño de la muestra y la estimación del poder estadístico sigue siendo inadecuado, independientemente del tipo de estudio¹⁴² y especialmente pobre en el campo de las afecciones musculoesqueléticas⁸⁴. La problemática involucra tanto la falta de un cálculo previo como el deficiente reporte del mismo⁸⁴. A efectos de esta tesis, consideramos un reporte adecuado aquel que incluyó los valores alfa y beta, el valor de la diferencia mínima clínicamente importante (MCID), la desviación estándar en variables continuas, la tasa -o media- esperada en el grupo control y el margen delta para estudios de no inferioridad. En el primer trabajo, detectamos que solo el 41% de los ECAs de nuestra muestra reportó un cálculo de la muestra¹³⁵ y, en el tercer trabajo, detectamos que el

36% reportó deficientemente este cálculo¹³⁶. Este deficiente reporte parece ser un hallazgo común en los ECAs sobre medicina física y rehabilitación¹⁴³, lo que limita la capacidad de los lectores para interpretar correctamente los resultados. Por ejemplo, en una muestra de 222 ECAs en el campo de la rehabilitación, *Castellini et al.* encontraron que sólo el 36% reportó el cálculo del tamaño de la muestra¹⁴². En la misma línea, *González et al.* llegó a la conclusión de que solo el 34% de los ensayos de fisioterapia relacionados con afecciones musculoesqueléticas reportaron un cálculo adecuado del tamaño de la muestra¹⁰². Por lo tanto, como primera consideración, se concluye que tanto la inclusión como el correcto y completo reporte del cálculo del tamaño de la muestra debe considerarse un requisito indispensable en los ECAs en TM. Sin embargo, existen otras consideraciones.

El cálculo del tamaño de la muestra implica elementos no solo científicos sino también éticos. Un ECA es legítimo siempre que el estudio tenga la potencia suficiente para responder a la pregunta de investigación. El cálculo del tamaño de la muestra y sus supuestos subyacentes juegan un papel central en el diseño y la justificación del estudio. Los ensayos no concluyentes con poca potencia podrían conducir a una exposición innecesaria de los participantes a riesgos y a una pérdida de recursos^{144,145}. Además, el reporte completo sobre cómo se calculó el tamaño de la muestra evita la dependencia exclusiva del poder estadístico y también aborda elementos clave como la importancia del resultado seleccionado y la MCID que se espera detectar^{146,147}. Centrarse demasiado en remarcar la significancia estadística puede enmascarar la importancia práctica o la relevancia clínica de un hallazgo concreto conduciendo a una interpretación simplista de los resultados. En nuestra muestra, solo el 41,6% de los ensayos especificaron la MCID, o bien explícitamente (30,8%) o bien como parte del cálculo del tamaño de la muestra (10,8%). Además, aunque 58,3% de los ensayos informaron de resultados estadísticamente significativos, solo el 32,5% reportaron resultados clínicamente relevantes. Debido a la naturaleza de las intervenciones en TM, los ensayos clínicos a menudo tienen como objetivo demostrar la efectividad mediante la aplicación de un enfoque pragmático al diseño del estudio. En este escenario, la MCID es imprescindible y proporciona una referencia valiosa para evaluar el impacto potencial de los hallazgos en la práctica clínica real¹⁴⁸.

Sin embargo, es importante mencionar que, si bien siempre es deseable contar con un tamaño de muestra adecuado, también hay que ponderar su relevancia con relación a otros elementos del diseño. Algunos autores han argumentado que los ensayos con escaso poder estadístico podrían ser aceptables si los investigadores utilizan el rigor metodológico para eliminar el sesgo, si reportan adecuadamente para evitar malas interpretaciones y si siempre publican los resultados para evitar el sesgo de publicación¹⁴⁹. Según *Schulz et al.*, los lectores deberían estar más preocupados por los errores sistemáticos (sesgos) que por un tamaño de muestra inadecuado. De hecho, el tamaño de la muestra se relaciona más con la precisión de la estimación del efecto y no necesariamente con la validez del estudio³¹. Es importante remarcar también que, el poder estadístico del ECA, se expresa en el intervalo de confianza. Por lo tanto, si se consideran cuidadosamente las cuestiones metodológicas, la potencia estadística podría no ser una preocupación importante¹⁴⁹. Contrariamente, la validez y la presencia de sesgos sí que puede influir drásticamente en el tamaño del efecto. Por ejemplo, en el campo del dolor lumbar, los estudios con mayor riesgo de sesgo reportaron tamaños del efecto que fueron, en promedio, un 50% mayores que las estimaciones reportadas en los ensayos con un menor riesgo de sesgo¹⁰⁴. En una muestra de ensayos que evaluaron tratamientos para la osteoartritis de cadera y rodilla, se demostró que otros elementos relacionados con la validez, como la generación inadecuada

cuada de la secuencia de asignación aleatoria, la falta de ocultación de la asignación y la ausencia de doble ciego, produjeron efectos de tratamiento más grandes⁷⁶.

5.2.2 Intervenciones control

En el campo de la medicina física, TM y rehabilitación, la escasez de investigación clínica experimental considerada de alta calidad podría explicarse parcialmente por las dificultades relacionadas con el uso de placebos o intervenciones simuladas (“sham”) como grupos control^{66,150}. Por ejemplo, en nuestra muestra, las limitaciones relacionadas con una intervención control inadecuada se reportaron en el 32,5% de los artículos revisados. Dentro de esta categoría, la razón principal declarada por los autores (41%) señaló que las intervenciones con placebo o simuladas no eran óptimas para la comparación¹³⁷. En el campo de las ICs (intervenciones con alta interacción entre pacientes y terapeutas), y muy especialmente en el caso de la TM, el efecto propiciado por el uso de placebos o *shams* está muy influenciado por el terapeuta que administra el tratamiento^{31,66}. Por lo tanto, como en todas la ICs, en los ECAs controlados con intervenciones placebos/sham manuales, aislar el “ingrediente activo” de otros efectos puede resultar muy complejo⁷¹.

Sin embargo, cabe la opción de preguntarse sobre la necesidad y/o idoneidad de aislar o no los efectos directos (específicos) de los indirectos (no-específicos) en los ECAs que evalúan los efectos de la TM. Autores relevantes en este campo conceptualizan la respuesta placebo como un mecanismo activo e importante de las TMs³⁰. Es sabido que en cualquier intervención basada en el contacto manual existe, en mayor o menor medida, un porcentaje del efecto que debe ser atribuido a la respuesta placebo-nocebo. Este porcentaje se añade a los efectos específicos de la intervención propiamente dicha y contribuye de manera decisiva en el efecto global de la intervención. Dentro de los mecanismos de acción descritos en la literatura sobre la TM, la respuesta placebo-nocebo se presenta como un elemento poderoso y siempre presente^{17,18}. Siendo este elemento indisociable a las intervenciones manuales, desde el punto de vista clínico se aboga claramente a estimular su potencial terapéutico⁸². Sin embargo, sigue siendo un tema muy controvertido en el contexto académico e investigador.

Mediante el uso de placebos apropiados es posible minimizar varios tipos de sesgo en los ECAs y mejorar la calidad de la investigación. En la literatura, existen algunas recomendaciones metodológicas en el campo de la rehabilitación y las TMs^{68,150,151}. Sin embargo, todavía no existe un placebo manual validado y de uso común⁷¹. De hecho, incluso entre los expertos en la materia, las creencias sobre lo que constituye un placebo apropiado en TM son muy diversas⁶⁸. En este escenario, cabe destacar, una vez más, la importancia de un reporte completo de la intervención de control. Por ejemplo, en una revisión de 64 estudios en el campo de la osteopatía, *Cerritelli et al.* encontró que los ECAs sistemáticamente reportaron de forma insuficiente los detalles sobre los procedimientos simulados, las dosis simuladas y detalles sobre el tratamiento simulado. Los autores concluyeron que estos hallazgos, junto con la alta heterogeneidad y la variabilidad encontrada entre los procedimientos simulados y los reales, comprometían profundamente la validez de dichos estudios⁷¹. Por lo tanto, las recomendaciones para futuras investigaciones deben incluir mejoras en los procedimientos y estrategias placebo/sham, a la vez que debe mejorarse el reporte tanto del grupo de intervención como control. La comunidad científica y clínica deben ponerse de acuerdo sobre lo que constituye la especificidad de un enfoque

manual¹⁵² para asegurarse de que el tratamiento simulado o placebo utilizado en los estudios no posee los mismos atributos que la intervención. Afortunadamente, con el objetivo de mejorar el reporte en estas intervenciones, recientemente se ha publicado una adaptación de la guía TIDieR para el reporte de controles placebo/sham¹⁵³.

5.2.3 Enmascaramiento

El enmascaramiento es un elemento crítico en todas las intervenciones no farmacológicas⁶⁷ y se considera un desafío importante en los ECAs de rehabilitación y medicina física^{31,50,53,150,154}. De hecho, la última edición de la guía CONSORTnpt (punto 11c) reconoce esta gran dificultad y recomienda, al menos, describir los intentos de limitar el sesgo cuando el enmascaramiento no es posible¹²². De hecho, al comparar ECAs que evalúan tratamientos no-farmacológicos versus farmacológicos, la falta de enmascaramiento en los artículos sobre tratamientos no-farmacológicos explica la mayoría de las diferencias metodológicas⁷⁶. El enmascaramiento se reporta con menos frecuencia en los ECAs que evalúan intervenciones no-farmacológicas, posiblemente debido a la dificultad para lograrlo y mantenerlo⁶⁷. Si bien el enmascaramiento de los evaluadores de los resultados debería ser factible en la mayoría de los casos, el enmascaramiento de los pacientes es un desafío y resulta casi imposible para los terapeutas manuales.

Los datos obtenidos en los trabajos de esta tesis doctoral confirman dichas dificultades en los ECAs sobre TM. En el primer trabajo evidenciamos que el enmascaramiento en los ensayos de TM es un aspecto que apenas ha mejorado a lo largo de los años, con una tendencia favorable solo evidente para el enmascaramiento del evaluador de resultados¹³⁵. En el segundo trabajo identificamos que las LARs asociadas a problemas con el enmascaramiento estaban presentes en casi el 30% de los estudios de la muestra (28,3%)¹³⁷.

Una de las investigadoras que más ha aportado a este campo es la Dra. Isabelle Boutron, actual directora del Centro Cochrane Francés, profesora de Epidemiología en la Universidad Paris Descartes y jefa del equipo INSERM-METHODS dentro del Centre of Research in Epidemiology and Statistics (CRESS). *Boutron et al.* publicó en 2007 una lista de métodos de enmascaramiento utilizados en ensayos no-farmacológicos y proporcionó directrices para superar algunas dificultades relacionadas⁶⁷. Entre otras posibles estrategias, propusieron cegar a los participantes en relación a la hipótesis de estudio (diseño Zelen modificado) y el uso de intervenciones de distracción de la atención o intervenciones de control con placebo no idénticas al tratamiento activo⁶⁷. Las fortalezas y limitaciones de cada método deben considerarse cuidadosamente en cada caso particular. Cabe también mencionar la reciente aportación realizada por el estudio meta-epidemiológico de *Moustgaard et al.*¹⁵⁵ en el que se concluye que tal vez el enmascaramiento podría ser menos importante de lo que se cree. En dicho estudio no se encontraron diferencias en el efecto del tratamiento estimado entre los ensayos con y sin enmascaramiento de pacientes, terapeutas o evaluadores de resultados. Sin embargo, los propios autores apuntan a posibles limitaciones de su propio diseño e invitan a seguir considerando el enmascaramiento como elemento indispensable del diseño de un ECA hasta que se confirmen sus resultados¹⁵⁵.

Es especialmente importante tener en cuenta que los ensayos con variables autoreportadas por los pacientes en ICs altamente dependientes del operador como la TM, pueden estar más fácilmente afectados por sesgos relacionados con las creencias y las ex-

pectativas^{156,157}. De hecho, la fuerza y la dirección de las creencias preconcebidas puede modificar potencialmente el efecto tanto del enmascaramiento¹⁵⁶ como de las propias intervenciones¹⁵⁸. En el campo de la afecciones musculoesqueléticas, esta influencia ha sido evidenciada en varias ocasiones^{157,159}. Por otro lado, hay que tener en cuenta que los efectos producidos por factores contextuales están bien documentados en la literatura sobre TM^{35,83,160,161} y pueden también tener un impacto sobre el enmascaramiento. En este escenario, obtener una medida *a priori* sobre la credibilidad del tratamiento y las expectativas de los participantes, además de registrar sistemáticamente las preferencias de los participantes al comienzo del ensayo (antes de la aleatorización), podría ayudar a ponderar los resultados y re-evaluar la magnitud del efecto de la intervención⁶⁷.

Por otro lado, el éxito en el enmascaramiento rara vez se evalúa y existe incertidumbre metodológica sobre cómo hacerlo¹⁶². Aunque, eventualmente, evaluar la eficacia del enmascaramiento podría considerarse una buena práctica a adoptar, se han reportado dificultades para registrar su impacto en relación a la prevención de sesgos^{163,164}. Por el contrario, se recomienda encarecidamente la evaluación previa de los procedimientos de enmascaramiento y el reporte explícito sobre quién fue cegado y a través de qué medios¹⁶⁵.

Tomando toda esta evidencia colectivamente y teniendo en cuenta las particularidades de la TM, consideramos que, aparte de asegurar un completo y detallado reporte, una buena estrategia para mejorar el enmascaramiento en este campo debe contemplar los siguientes elementos: a) la elección de controles placebo/sham lo más validados posibles⁶⁶; b) el análisis previo y exhaustivo sobre la estrategia de enmascaramiento a implementar (valorando fortalezas, limitaciones e implicaciones éticas)⁶⁷; c) la obtención de un registro previo sobre las preferencias de los participantes en relación a las intervenciones (especialmente en ECAs con diseño abierto)^{158,159} y d) el diseño de estrategias de control de la eficacia de dicho enmascaramiento¹⁶⁶.

5.2.4 Duración del estudio y seguimiento

Colectivamente, los datos científicos sobre los mecanismos específicos de la TM informan que los efectos principales de dicho abordaje son a corto plazo^{17,18}. Sin embargo, como se ha comentado anteriormente, en la práctica clínica, la TM se utiliza habitualmente para tratar enfermedades crónicas¹³⁸⁻¹⁴⁰ mediante paquetes terapéuticos que pueden durar varias semanas o meses. Aparte de las consideraciones clínicas que se derivan de este hecho, existen también implicaciones relacionadas con la investigación en este campo. En primer lugar, parece lógico fomentar el seguimiento a largo plazo en los ensayos clínicos de TM que evalúen condiciones crónicas. De hecho, idealmente, los investigadores de TM deberían desarrollar modelos de intervención coincidentes con la historia natural de las patologías de estudio, considerando la evolución de los síntomas a lo largo del tiempo como un resultado en lugar de establecer periodos de seguimiento arbitrarios¹⁶⁷. Sin embargo, el elemento del tiempo y sus condicionantes afecta a los ECAs en TM. En nuestra muestra, solo el 39% de los ECAs reportaron resultados a largo plazo (>3 meses) y la duración de los estudios y su seguimiento fue una limitación común reportada en el 33,3% del total de los casos. Los principales motivos para declarar esas limitaciones se relacionaron con la falta de seguimiento a largo plazo (87,5%).

Por otro lado, la evaluación completa de todos los participantes incluidos en un ECA es crucial para una interpretación sólida de los resultados¹⁶⁸. Como sucede habitualmente, varios participantes inicialmente aleatorizados se perderán durante el seguimiento o no se analizarán por algún motivo al final del estudio, lo que provocará un potencial sesgo de desgaste. Los datos no obtenidos debido a la negativa de los participantes a completar las evaluaciones de resultados, a las visitas no realizadas o a la pérdida del seguimiento del tratamiento y sus evaluaciones, pueden afectar el poder del estudio¹⁶⁹. Por tanto, el reporte preciso de todos estos acontecimientos es fundamental para una correcta evaluación de los resultados. En el primer trabajo de esta tesis identificamos un buen reporte general de estos elementos y, además, se demostró un cambio significativo en la inclusión del diagrama de flujo entre el grupo pre-C y el grupo post-C¹³⁵. Del mismo modo, en el tercer trabajo identificamos que el 83,3% de los artículos analizados reportó suficiente información para evaluar las pérdidas durante el seguimiento¹³⁶. En conjunto, se concluye que el número de participantes asignados al azar y la información sobre los datos faltantes (*missings*), los abandonos y las pérdidas durante el seguimiento generalmente se reportan bien en los ensayos de TM.

Más allá de la inclusión de un diagrama de flujo, las recomendaciones efectuadas en las guías y extensiones CONSORT incluyen la necesidad de especificar el número de pacientes perdidos durante el seguimiento pero también los motivos de dichas pérdidas y el número de pacientes finalmente analizados al final del estudio para cada variable de resultado^{108,170}. Para una correcta interpretación de los resultados, los autores deben indicar claramente cómo fue la gestión de los datos faltantes y el modelo estadístico utilizado para analizar el ensayo. A este respecto, CONSORT apela a reportar directamente si el análisis fue o no realizado por intención de tratar (ITT) (*“stating the number of participants included in each analysis and whether the analysis was by intention-to-treat”*). Consideramos que este aspecto es especialmente relevante en el campo de la TM. Dadas sus características intrínsecas, los ECAs en TM a menudo tienen como objetivo evaluar la efectividad (más que la eficacia) de las intervenciones, y como se ha comentado, para ello en muchas ocasiones se aplica un enfoque pragmático al diseño. Se ha sugerido que el análisis de los resultados debería aplicar el mismo enfoque pragmático, es decir, reflejar la práctica clínica de la forma más fiel posible. El ITT ayuda a este propósito proporcionando una indicación más realista de los beneficios clínicos del tratamiento¹⁷¹.

En el primer trabajo de este compendio (n=100) detectamos que los participantes incluidos en el análisis realizado en los ECAs fueron reportados en el 97% de los casos. Sin embargo, solo el 34% de los artículos se refirió al ITT y la imputación de los datos faltantes se describió en solo la mitad del 66% de los artículos que reportaron tasas de abandono¹³⁵. En el tercer trabajo (n=120), detectamos que en el 83,3% de los artículos de la muestra se aportó suficiente información para evaluar si se produjo alguna pérdida de seguimiento. Sin embargo, al analizar detalladamente esa información, encontramos que los motivos de la pérdidas no se reportaron en el 23% de los casos y los criterios de imputación de los datos faltantes en el análisis no se describió en el 22% de los casos¹³⁶. La inclusión u omisión de estos elementos (las razones para las pérdidas y los criterios de imputación aplicados a estas pérdidas) en el reporte de un ECA permite entender qué tipo de ITT se ha realizado y puede afectar directamente a la valoración de los resultados del estudio.

Sin embargo, sabemos que un reporte completo sobre el ITT se incluye con poca frecuencia en las publicaciones sobre terapias físicas¹¹⁰. Los ensayos pueden no cumplir con estos criterios por varios motivos: por no utilizar el ITT, por no informar que se utilizó el análisis

ITT o por malinterpretar la definición del ITT. Hay que tener en cuenta también que incluir un seguimiento completo de los participantes y/o el hecho de tener dos grupos iguales al final del estudio no garantizan la calidad del ITT¹⁷². Hasta la fecha, ninguno de los artículos originales sobre el ITT, el Manual Cochrane, la declaración CONSORT o la Escala PEDro, recomiendan específicamente incluir esta información¹⁷². Sin embargo, para favorecer una mayor transparencia en la interpretación de los datos, sugerimos que se recomiende la inclusión de una declaración formal sobre el tipo de análisis efectuado en el manuscrito publicado. A modo de soporte, la revisión publicada por *Armijo-Olivo et al.* ofrece una guía muy útil para comprender cómo lidiar con los datos faltantes en la investigación clínica¹⁶⁹.

5.2.5 Intervención experimental

Tal como se ha explicado en la introducción de esta tesis, cualquier IC implica en su ejecución la suma de múltiples elementos que impactarán en mayor o menor grado en el resultado. Sin embargo, dentro de este tipo de intervenciones, la TM constituye un caso paradigmático por el hecho de utilizar el tacto como elemento central. La compleja interacción paciente-terapeuta que se establece cuando se toca a un sujeto con intención terapéutica convierte la TM en una forma específica de contacto humano¹⁶¹. La estimulación táctil no solo impacta en las expectativas del paciente sino que representa una herramienta especial para comunicar de manera no verbal mensajes significativos al cerebro del paciente. El tacto es capaz de producir analgesia, regular las emociones del paciente y reorganizar las representaciones mentales¹⁶¹. Sin embargo, todos estos procesos suceden en distinto grado y de forma individual en cada una de las interacciones terapéuticas que se dan, siendo además altamente dependientes del contexto y de las características de los protagonistas de dicho encuentro terapéutico (pacientes y terapeutas). Por este motivo, la idoneidad o no de protocolizar y/o estandarizar las intervenciones manuales suscita diversidad de opiniones en el seno del colectivo y adquiere perspectivas diferentes en función de si se mira bajo el prisma clínico o investigador.

Si bien es razonable afirmar que, en la práctica clínica diaria, la individualización de las intervenciones manuales es el proceder habitual, en el campo de la investigación este elemento es controvertido. Los requerimientos metodológicos establecidos en la valoración de la validez interna de los estudios fuerza a los investigadores en el campo de la TM a reducir el grado de individualización a favor de protocolos estandarizables y reproducibles. De hecho, algunos autores afirman que la estricta supeditación a tratamientos basados en clasificación, reglas de predicción o guías de práctica clínica puede explicar parcialmente la falta de evidencia sobre los efectos a largo plazo de la TM²⁰. Como alternativa, proponen un acercamiento a la realidad clínica a través de la aplicación de tratamientos basados en un proceso de razonamiento clínico en el que se considere de forma individual la evolución del paciente y los cambios entre sesiones a lo largo de todo el estudio²⁰. En algunos casos, es posible trabajar con algoritmos o flujos de trabajo, donde los pasos en la ruta de tratamiento dependen del resultado del paso anterior^{20,31,55}. De la misma manera, en TM, la elección del tratamiento suele depender de los resultados de una evaluación analítica exhaustiva al comienzo de cada sesión¹⁷³. Este proceso de evaluación analítica permite a los terapeutas basar sus opciones de tratamiento en el historial, los signos y los síntomas del paciente y, de esta manera, el proceso puede documentarse. Si la elección de la intervención ocurre inmediatamente después de una evaluación realizada como parte del proceso de razonamiento clínico⁵⁵, estandarizar dicha evaluación y establecer un tratamiento

determinado en función de estos resultados ayuda a incrementar la reproducibilidad de la intervención a la vez que respeta la aplicación individual de la misma.

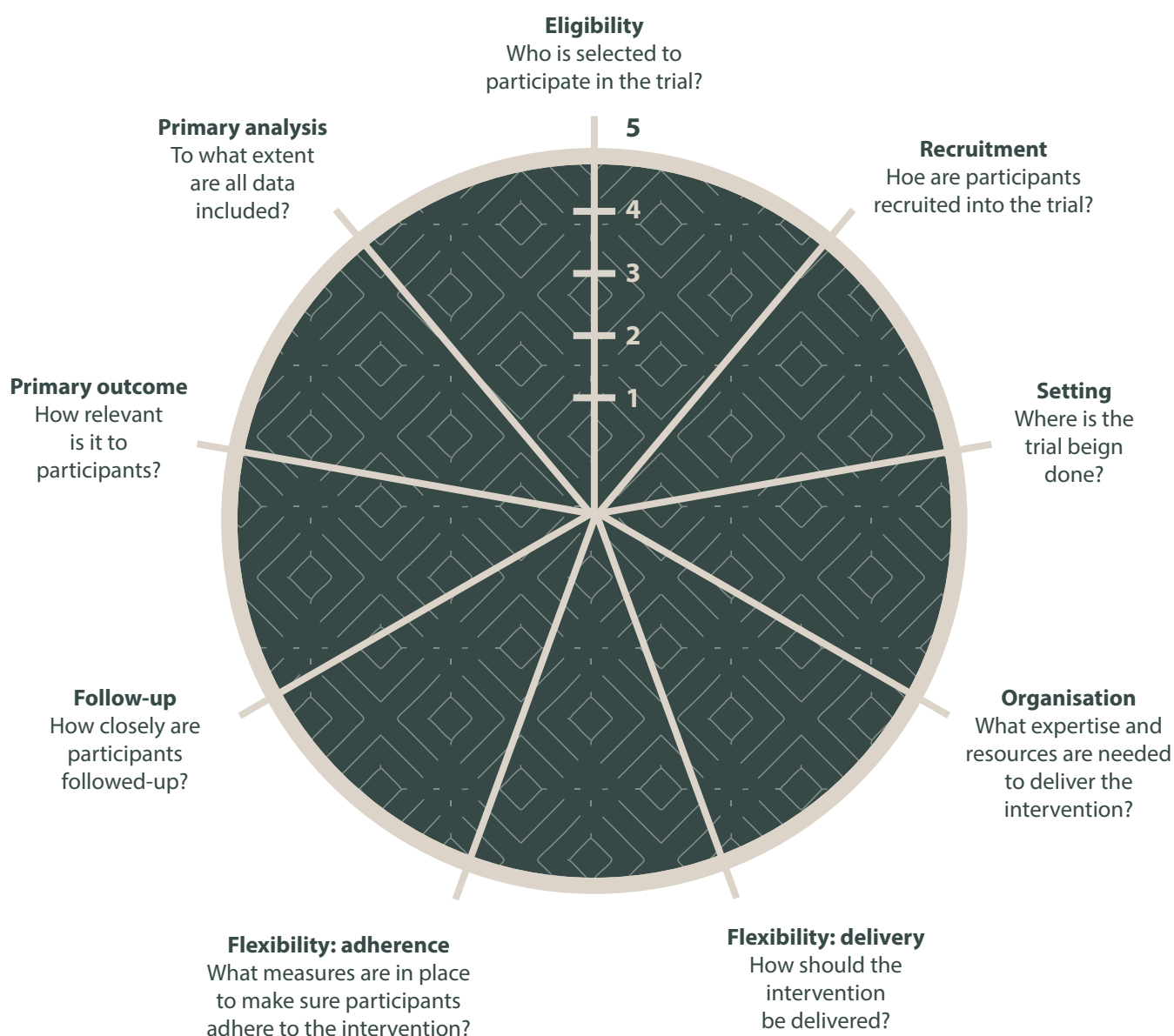
El posicionamiento sobre la importancia del tratamiento individualizado se refuerza aún más al considerar la naturaleza multidimensional de muchas de las patologías comúnmente tratadas por los terapeutas manuales, como es, por ejemplo, el dolor lumbar¹⁷⁴. Parece poco coherente aplicar intervenciones estandarizadas para tratar patologías con etiología multifactorial. Sin embargo, si bien existen algunos ECAs claramente diseñados bajo la perspectiva de la individualización^{94,98,175,176}, también existen estudios en los que se cuestiona su supuesto beneficio en el campo de la TM¹⁷⁷.

En el segundo de los trabajos del compendio que conforma esta tesis, identificamos un 27,5% de ECAs con LARs relacionadas con la intervención experimental. Dentro de este subgrupo, la principal razón reportada por los autores para reconocer esta limitación (51,5% de los casos) apuntó a la falta de un protocolo estandarizado para la intervención experimental¹³⁷. Curiosamente, este dato contrasta con el hecho de que otro 15,2% reportaron exactamente lo contrario, es decir, la falta de individualización del tratamiento como una limitación específica. Estos hallazgos confirman la doble tendencia existente dentro de este campo consistente en el hecho de que algunos investigadores apuestan por intervenciones protocolizadas que favorezcan altos niveles de validez interna frente a otros que apuestan por un enfoque pragmático hacia las intervenciones basado en la individualización de los procedimientos, tal y como se hace corrientemente en la práctica clínica²⁰. En las disciplinas más holísticas de la TM, este debate se extiende a si las intervenciones deben realizarse bajo una perspectiva corporal global o bajo un enfoque local y/o sintomático. Una vez más, los requerimientos metodológicos favorecen el uso de intervenciones lo más concretas y homogéneas posible basadas en criterios diagnósticos verificables. Esto empuja a los investigadores a restringir abordajes más globales (y probablemente más subjetivos) comúnmente utilizados en la práctica clínica en favor de técnicas específicas dirigidas a áreas corporales sintomáticas. Por ejemplo, en nuestra muestra, el 24,2% de las LARs relacionadas con la intervención experimental se justificaron porque el tratamiento se administró solo en una determinada zona del cuerpo y no bajo una perspectiva global.

De todo lo mencionado anteriormente, se concluye que las particularidades intrínsecas de las intervenciones basadas en el tacto y centradas en el paciente comprometen de una manera muy clara el equilibrio entre la validez interna y externa de los ECAs. Concretamente, en relación a las intervenciones experimentales, está en la mano de los investigadores la decisión de optar por un enfoque más explicativo y, por lo tanto, diseñar intervenciones más homogéneas, reproducibles y estandarizadas o, por el contrario, optar por un enfoque más pragmático con un nivel más alto de individualización y flexibilidad. De hecho, esta decisión trasciende la elección del enfoque terapéutico y es aplicable a otros elementos del diseño del ECA. Por ejemplo, la presencia de unos criterios de inclusión estrictos favorecerá, sin duda, la obtención de una muestra homogénea. Sin embargo, teniendo en cuenta que la población comúnmente tratada por los terapeutas manuales presenta numerosas comorbilidades¹⁷⁸⁻¹⁸⁰, probablemente una inclusión de pacientes restrictiva impactará dramáticamente en la generabilidad de los resultados¹³⁷. A este respecto, conviene recordar la propuesta de *Schwartz y Lellouch* realizada en 1967 sobre la existencia de un continuo pragmático-explicativo en el diseño de los ECAs¹⁸¹. Tal y como fue definido originalmente, los dos extremos del continuo deben ser interpretados más

como una actitud concreta con relación al diseño de los ECAs y no como dos diseños diferenciados. Esta idea quedó perfectamente representada en el desarrollo original de la herramienta *Pragmatic Explanatory Continuum Indicator Summary* (PRECIS) en el 2009¹⁸² y en su posterior actualización en 2015¹⁸³. Esta herramienta, ideada para dar soporte al diseño de los ensayos, constituye un claro indicador sobre la actitud del mismo en relación a 9 dominios esenciales del diseño de un ECA. Entre ellos, uno de los más determinantes se refiere a la forma en cómo se aplican las intervenciones (Dominio 5) **Fig 5**.

Fig 5: PRECIS Wheel. (Adaptado de Loudon et al. 2015)



Como IC, la evaluación de la TM precisa tanto de estudios experimentales de corte explicativo como de estudios de corte pragmático. Los primeros contribuyen a entender los mecanismos de acción de la TM, sus bases fisiológicas y la eficacia de las técnicas manuales aplicadas aisladamente. Los estudios de corte pragmático acercan la TM a la práctica clínica habitual aportando datos sobre la efectividad de los tratamientos, la eficiencia, los condicionantes contextuales y los efectos específicos. Ambos enfoques poseen fortalezas y debilidades¹⁸⁴ y por ello se precisa de la suma de varios diseños experimentales más la evidencia aportada por estudios observacionales para obtener una valoración completa de una IC como la TM^{49,95}.

Sin embargo, más allá del diseño y la actitud del ECA, uno de los elementos más importantes con relación a las intervenciones (tanto experimentales como los controles) es su adecuado reporte y descripción. El informe detallado de cualquier intervención es crucial para evaluar su aplicabilidad al contexto clínico y evitar un desperdicio de recursos^{185,186}. Sin embargo, se ha estimado que, en cualquier ámbito médico, hasta el 60% de la literatura publicada reporta de forma incompleta las intervenciones¹⁸⁷. Este problema, común para todos los tipos de tratamientos e intervenciones, es significativamente peor en los ECAs no-farmacológicos en los que tan solo el 29%-39% de las intervenciones se describen adecuadamente en comparación con el 67% de las intervenciones farmacológicas^{187,188}. Esta deficiencia en el reporte de las intervenciones es común en el campo de las terapias físicas¹⁸⁹, incluyendo la TM¹⁰².

Los resultados del primer trabajo de esta tesis mostraron que, a pesar de algunas mejoras (en el grupo post-C) en la descripciones de los participantes, los terapeutas y el ámbito de realización del estudio (*setting*), todavía es necesario proporcionar descripciones más detalladas, en especial respecto a las intervenciones¹³⁵. Entre las diversas iniciativas que se han emprendido para mejorar la descripción de las intervenciones, la guía TIDieR¹⁹⁰ debe destacarse como la más prometedora. De hecho, como trabajo adicional al compendio que conforma esta tesis doctoral, adjuntamos un artículo publicado en 2016 en el que señalamos la importancia de esta guía para la descripción de las intervenciones en TM¹²⁴ (**Anexo 4**). La Guía TIDieR supone una extensión del ítem 5 de la guía CONSORT¹⁰⁸ y del ítem 11 de la guía SPIRIT (*Standard Protocol Items: Recommendations for Interventional Trials*)¹⁹¹. Gran parte de su contenido se ha incorporado en la actualización de la extensión CONSORTnpt realizada en 2017¹²² y recientemente se ha publicado una versión de la misma guía específicamente orientada a intervenciones placebo¹⁵³. Sin lugar a dudas, el uso de la guía TIDieR debe considerarse en el reporte de los ECAs en TM, aunque su utilización en el diseño de los mismos es también recomendable. De hecho, el debate sobre la individualización de las intervenciones se aborda en el ítem 9 (*"Tailoring"*). Con la inclusión de este ítem, la guía TIDieR hace un avance significativo en comparación con otras guías, reconociendo las peculiaridades de algunas intervenciones en las que la adaptación al paciente y la flexibilidad son relevantes¹²⁴.

5.2.6 Medidas de resultado

La producción científica en TM ha aumentado considerablemente en las últimas dos décadas. Sin embargo, la evidencia sobre la efectividad general de las intervenciones manuales sigue siendo bastante controvertida²⁰. Varias revisiones sistemáticas recientes¹⁹²⁻²⁰⁰ muestran que los resultados de los ECAs en TM no son concluyentes, lo que plantea la necesidad de realizar estudios con una mejor metodología y reporte²⁰¹⁻²⁰⁶. Un reporte completo permite la interpretación óptima de los métodos y especialmente de los resultados, lo que facilita la evaluación del efecto de la intervención y sus implicaciones para la práctica clínica^{207,208}. De hecho, las limitaciones en el reporte de los resultados están asociadas con estimaciones sesgadas de los efectos del tratamiento²⁰⁹. En el tercer trabajo del compendio que forma esta tesis doctoral realizamos un análisis exhaustivo de la calidad del reporte de los resultados. Nuestros hallazgos demostraron que el reporte de los resultados en los ECA en TM es deficiente¹³⁶. Entre los elementos que analizamos es conveniente discutir algunos por su relevancia en la interpretación de los resultados.

A pesar de que las guías CONSORT recomiendan que el resultado primario debe reportarse específicamente y de forma diferenciada de los resultados secundarios^{108,209}, en nuestra muestra menos del 50% de los artículos declararon explícitamente el resultado primario del estudio, aunque éste fue deducible a partir del cálculo del tamaño de la muestra en el 7% de los casos restantes. La medida de resultado primaria se establece en el momento en que se diseña el estudio y se redacta el protocolo. Los resultados se basan directamente en esa variable primaria²¹⁰ y el tamaño de la muestra se calcula también a partir de ella. Por lo tanto, declarar claramente el resultado o variable primaria evita que los investigadores seleccionen a posteriori otros hallazgos y que los presenten como los principales del estudio. Además, para evitar este sesgo de reporte, se recomienda encarecidamente registrar de forma prospectiva el protocolo del estudio en un registro público de ensayos.

Aunque un ECA tiene como objetivo comparar el efecto entre dos intervenciones, en algunos casos, la evaluación de los efectos dentro de cada grupo también puede dar una información valiosa. Especialmente cuando las intervenciones muestran beneficios clínicos pero la comparación entre grupos no alcanza significación estadística. Por esta razón, el reporte del tamaño del efecto y la precisión de este, tanto dentro (intra) como entre grupos, es importante para interpretar correctamente los resultados y proporcionar una comprensión amplia del efecto de una intervención. Todos los ECAs de nuestra muestra reportaron efectos intra-grupo (antes versus después), y de hecho, en el primer trabajo de esta tesis evidenciamos una mejora significativa del reporte en este aspecto en el grupo post-C¹³⁵. Sin embargo, solo el 46,7% de los ECAs reportó con detalle el tamaño del efecto y su precisión en la comparación entre grupos¹³⁶. Esta falta de información plantea importantes consideraciones clínicas en campos como la TM, donde los efectos inespecíficos juegan un papel importante en el efecto general^{18,35,211}. Mientras que el valor intra-grupo informa sobre los efectos del tratamiento, este valor también incluye aspectos como la recuperación natural, la regresión a la media o efectos no específicos. Sin embargo, la valoración entre-grupos ofrece una estimación objetiva del efecto del tratamiento²¹². Por lo tanto, el cálculo y la inclusión de ambas estimaciones debe considerarse una práctica recomendada en el análisis estadístico y en el reporte de ECAs en TM.

Otro aspecto importante se relaciona con las medidas de precisión. La mitad de nuestra muestra (47,5%) solo reportó el valor *p* para la comparación entre grupos. Si analizamos

las medidas de precisión para todas las estimaciones realizadas, el valor p se utilizó en el 88,3% de los casos, pero el intervalo de confianza se reportó solo en el 29,2%¹³⁶. Aunque los valores de p se calculan fácilmente, a menudo dan lugar a malentendidos^{147,213,214}. Además, a menudo también se utilizan de forma incorrecta²¹⁵ y pueden sobreestimar los efectos del tratamiento²¹⁶. El uso de intervalos de confianza como medida de precisión de la estimación de los efectos conduce a una mejor interpretación clínica de la literatura y ofrece mayor fiabilidad para analizar la magnitud real del efecto observado^{147,217}. Con el fin de mejorar la interpretación y los informes estadísticos a este respecto, cabe destacar la guía publicada por *Greenland et al.* en el 2016²¹⁴.

En nuestra muestra, el 29,2% de las LARs se relacionaron con las medidas de resultado. Entre las razones dadas por los autores, la falta de datos y la validez de las mediciones fueron las más frecuentes (34,3% y 31,4%)¹³⁷. La selección de medidas de resultado válidas y clínicamente significativas es crucial para que la investigación sea transferible a la práctica clínica. Además, debe tenerse en cuenta que las herramientas para la medición de los resultados utilizadas en la investigación pueden no siempre ser las más apropiadas en contextos clínicos²¹⁸. Las dificultades causadas por la heterogeneidad en las mediciones de los resultados se reportan constantemente en las revisiones sistemáticas²¹⁹ y constituyen un desafío en el metanálisis de las intervenciones de fisioterapia²²⁰. El uso de un paquete de medidas de resultado estandarizadas para una condición determinada (compuesto de diferentes tipos de variables) sería una posible solución a estos problemas. A este respecto, la iniciativa COMET (*Core Outcome Measures in Effectiveness Trials* - <http://www.comet-initiative.org>) representa una herramienta muy útil y recomendable para los investigadores de TM. Como recomendación adicional, el uso de medidas de resultado informadas por el paciente (*Patient-reported Outcome Measures* - PROMs) constituye una opción adecuada para intervenciones centradas en el paciente como la TM. De hecho, las PROMs se analizan cada vez más en la investigación en fisioterapia y también son útiles para su uso en entornos clínicos²²¹. Una extensión CONSORT específica centrada en PROMs²²² proporciona pautas a los investigadores para asegurar un reporte óptimo.

Los datos reportados en una publicación incluyen tanto la expresión numérica de los resultados en el texto como su representación en gráficos o figuras. Las figuras juegan un papel clave a la hora de informar sobre los hallazgos y facilitar al lector su interpretación. Esto los convierte en valiosas herramientas de uso común para presentar datos para la comunicación científica²²³. El impacto visual ayuda a transmitir los hallazgos al lector²²⁴, ayudando en su interpretación y a tener una impresión duradera de los resultados²²³. *Puhan et al.*, en 2006²²⁵ definió tres requisitos principales para una comprensión óptima de un gráfico: i) primero, que los símbolos relacionados con datos deben ser claramente distinguibles para permitir la detección visual; ii) segundo, que los datos deben estar organizados de tal manera que se facilite la estimación de los valores importantes y sus relaciones; iii) y tercero, que debe haber una explicación completa de todos los elementos del gráfico y de cómo se obtuvieron los datos. Sin embargo, la presentación o el contenido de muchas figuras publicadas es deficiente²²⁶ convirtiéndose en una de las principales deficiencias de los ECAs^{227,228}. En otro estudio, *Pocock et al.*²²⁶ observaron que los principales tipos de figuras utilizadas en los ensayos clínicos son: diagramas de flujo, diagramas de Kaplan Meier, diagramas de Forest y mediciones repetidas a lo largo del tiempo. No es sorprendente, pues, que el tipo de gráfico más común identificado en nuestra muestra fue el gráfico de medidas repetidas. Esto se debe a que las variables continuas son las más utilizadas en los estudios en TM.

Nuestros resultados en relación a la representación gráfica de los resultados coincidieron con las deficiencias ya encontradas en otros campo de la salud¹³⁶. De hecho, las áreas con un reporte gráfico deficiente son similares independientemente del campo de estudio¹³⁴. Nuestros resultados representan un primer acercamiento a la evaluación de estas características en el campo de la TM y apuntan a la necesidad de mejorar en este aspecto. Para ello, y aunque este campo está poco explorado, es recomendable seguir algunas guías existentes como las publicadas por *Chen et al.* en el 2017¹³⁴.

Tabla 3: Deficiencias metodológicas identificadas y propuesta de mejora/solución

Elemento metodológico	Deficiencia identificada	Propuesta de mejora y/o solución
Tamaño de la muestra	• Tamaños pequeños • Falta de cálculo previo	• Inclusión y reporte adecuado del cálculo del tamaño de la muestra • Especificación del Valor Mínimo de Diferencia Clínica Importante a detectar • Con muestras pequeñas, extremar el rigor en la aplicación de los métodos y en la transparencia del reporte
Intervenciones control	• Ausencia de placebo/sham validado	• Mejorar las estrategias y procedimientos placebo/sham • Aumentar la precisión de reporte para dichos controles • Uso de TIDieR-Placebo/Sham como guía de referencia
Enmascaramiento	• Enmascaramiento de terapeutas y pacientes	• Asegurar el simple ciego enmascarando los evaluadores • Obtener una medida a priori sobre la credibilidad del tratamiento • Valorar expectativas y obtener un registro de preferencias de los pacientes • Valorar métodos como el enmascaramiento de los participantes en relación a la hipótesis de estudio (Zelen modificado) o el uso de placebos no parecidos a la intervención experimental • Diseñar e implementar estrategias de control del éxito del enmascaramiento • Descripción y reporte preciso de las estrategias adoptadas para minimizar el sesgo de detección cuando el enmascaramiento no ha sido posible
Duración el estudio y seguimiento	• Falta de seguimientos a largo plazo • Reporte incompleto de las pérdidas y su gestión	• Reportar adecuadamente las pérdidas (especialmente los motivos) • Priorizar el análisis por ITT y reportar de forma precisa el tipo de análisis efectuado y la gestión estadística de los datos faltantes (asunciones, imputaciones)

Intervención experimental	● Falta de protocolos estandarizados ● Presencia de protocolos estandarizados ● Reporte incompleto de las intervenciones	● En la evaluación de técnicas: Aplicar protocolos reproducibles ● En la evaluación de tratamientos: Establecer marcos, algoritmos o flujos de trabajo en el que los pasos en la ruta de tratamiento dependen de una evaluación previa (estandarizable) o del resultado del paso anterior. ● Uso de la herramienta PRECIS en la fase de protocolo (especialmente dominio 5) para determinar la actitud del ECA en el continuo explicativo-pragmático. ● Uso de la herramienta TIDieR (y extensiones) para el diseño y correcta descripción de las intervenciones.
Medidas de resultado	● Reporte deficiente de los resultados (tanto en el texto como gráfico)	● Uso de guía InsPECT (en desarrollo) ● Reportar claramente la variable primaria diferenciado del resto de variables de interés ● Registro del protocolo de estudio en un registro público de ensayos ● Reportar de forma clara y completa (efecto y precisión) lo efectos inter-grupo además de los intra-grupo ● Inclusión de intervalos de confianza para cada estimación ● Uso de paquetes de medidas de resultado estandarizadas (iniciativa COMET) ● Uso de medidas de resultado informadas por el paciente (PROMs) ● Asegurar un reporte gráfico de los resultados adecuados (tipo de gráfico adecuado, completitud de los datos y claridad visual)

5.3 Las guías de reporte

La obtención de los datos para la realización de los trabajos que conforman esta tesis doctoral (especialmente los obtenidos en el primer trabajo) se ha basado, en gran parte, en el análisis del reporte de las publicaciones científicas en TM. Estando nuestra muestra formada exclusivamente por ECAs, la guía de reporte de referencia fue la guía CONSORT y, en concreto, la extensión CONSORT para intervenciones no-farmacológicas en su primera versión de 2008¹²⁰. El análisis se complementó con el uso de la guía TIDieR¹²³, aunque su impacto solo se ha podido analizar parcialmente dada su fecha de publicación en el 2014. Es oportuno aclarar que nuestra muestra de artículos se extrajo de la literatura científica sin tener en cuenta si las revistas habían especificado su adherencia a la declaración CONSORT o no. Aún así, nuestros resultados fueron muy similares a otras publicaciones realizadas en el campo de la TM cuyo análisis se focalizó únicamente en artículos extraídos de revistas adherentes con la declaración CONSORT¹⁰⁰. De hecho, un estudio similar realizado en el campo de la quiropráctica²²⁹ exploró las asociaciones entre ciertas variables y la calidad general del reporte. Los autores encontraron que el reporte estaba influenciado por el año de publicación y el tamaño de la muestra, pero no por el tipo de revista, la fuente de financiación o los resultados positivos²²⁹. El uso concienzudo y sistemático de las guías de reporte durante la realización de esta tesis doctoral nos ha permitido observar tanto sus virtudes como sus limitaciones, y es por este motivo que consideramos interesante aportar nuestras reflexiones al respecto en el contexto de su aplicación en el campo de la TM.

Como primera reflexión, vemos que, si bien nuestros resultados acerca del reporte deficiente son consistentes con otras revisiones similares en el campo de las terapias físicas^{100,102}, estos contrastan con los resultados positivos encontrados en otros campos^{47,111,113,230}. Esto podría sugerir que la aplicación de las guías CONSORT puede tener un impacto diferente según el campo de estudio. Por otro lado, aunque es de esperar que exista una asociación entre la baja calidad metodológica de un estudio y su reporte deficiente, los resultados hasta la fecha y los comentarios realizados por varios autores plantean algunas cuestiones específicas sobre el uso de las guías de reporte y su eficacia para mejorar la literatura científica. Seguidamente planteamos algunas de estas preguntas.

5.3.1 ¿Son los elementos de las guías CONSORT suficientemente claros para los autores de manuscritos?

La existencia de recursos y el acceso a directrices para comprender cómo realizar un buen reporte de los manuscritos ha sido recientemente fuente de discusión^{119,231}. Además, se constata que las instrucciones específicas sobre cómo los autores deben utilizar CONSORT no son coherentes entre las revistas y editoriales¹¹⁸. Hasta donde sabemos, a día de hoy no existen encuestas realizadas a los autores con respecto a la claridad de los elementos de la lista de verificación CONSORT. Sin embargo, los estudios que evalúan estrategias enfocadas a ayudar a los autores a preparar manuscritos más completos (por ejemplo, Web-CONSORT: <https://www.webconsort.com>) concluyen que se necesita información más detallada sobre cómo implementar cada elemento en el contexto de un ensayo específico²³². A modo de ejemplo, la experiencia acumulada en la realización de esta tesis nos permite realizar algunas consideraciones. Nuestros trabajos fueron llevados a cabo por un equipo

de revisión compuesto por expertos clínicos en TM e investigadores con amplia experiencia en los aspectos metodológicos de los ECA. Si bien la prueba piloto realizada con anterioridad al análisis dio como resultado un acuerdo sustancial entre los revisores, durante el proceso surgieron claras diferencias de criterio que tuvieron que resolverse mediante discusión. De hecho, el trabajo de conformar un acuerdo final entre los pares después de una primera revisión individual ocupó tanto o más tiempo que la revisión original a pesar del nivel de acuerdo previo registrado en el estudio piloto. A la luz de estas dificultades, una primera observación que hacemos es que tanto el uso de un enfoque estructurado para reportar las investigaciones realizadas²³³ como el uso de herramientas de ayuda para la redacción como COBWEB (<https://cobweb.clinicalepidemio.fr>)²³⁴ deben ser estrategias consideradas para ayudar a los autores a mejorar la adherencia a las directrices de presentación de informes en la investigación sanitaria²³¹.

Otra pregunta que cabe hacerse es si existe un proceso estandarizado para evaluar la adherencia de los artículos a las guías de reporte. Más allá de las recomendaciones generales, consideramos que debe implementarse un mecanismo que asegure que la presentación y aceptación de artículos está condicionada a la inclusión de la información requerida en las guías de reporte^{118,235}. Además, los investigadores y revisores deberían recibir formación sobre el uso de estas directrices^{235,236}, y los editores deberían incorporarlas en el proceso editorial, ya que existen inconsistencias entre lo que los autores afirman en las listas de verificación enviadas y lo que realmente se reporta en el artículo publicado²³⁵. Este requisito no solo mejoraría la calidad de los informes sino que también podría mejorar potencialmente la calidad metodológica de los estudios.

5.3.2 ¿Es posible que la utilidad de las guías CONSORT difiera dependiendo del campo de estudio?, ¿Es la extensión para intervenciones no-farmacológicas adecuada para todas las intervenciones no farmacológicas?

Hasta la fecha, el número de estudios que apuntan a un impacto significativo de las directrices CONSORT (o sus extensiones) pertenecen mayoritariamente al dominio farmacológico^{111,230,237} en comparación con el no-farmacológico^{114,116,187,188}. La denominación “no-farmacológica” constituye una etiqueta genérica que incluye un gran número de intervenciones cuya naturaleza puede ser muy distinta. Parece razonable considerar que el mismo conjunto de directrices para optimizar el reporte puede no cubrir adecuadamente las características específicas de diferentes tipos de intervenciones. De hecho, investigaciones recientes muestran que, a pesar de la existencia de estas guías, la reproducibilidad clínica se ve comprometida particularmente para las intervenciones complejas²³⁸. Opinamos que la existencia de guías específicas según campos sanitarios particulares dentro de las llamadas intervenciones no-farmacológicas podría mejorar el reporte por dos motivos. En primer lugar, los elementos fundamentales para la calidad del reporte serían adaptados a las particularidades de cada disciplina. En segundo lugar, en un momento donde el número de guías disponibles es excesivo, los investigadores podrían identificar fácilmente la guía de reporte indicada para cada especialidad, potencialmente mejorando su adherencia. Por ejemplo, más allá de los resultados ya mencionados en el campo de las terapias físicas, la guía CONSORT ha tenido un impacto limitado en la mejora de la calidad del reporte en los campos (no-farmacológicos) de la Odontología y la Cirugía^{114,239}. Sin em-

bargo, la guía *Standards for Reporting Interventions in Clinical Trials of Acupuncture* (STRIC-TA)²⁴⁰ ha tenido un impacto generalmente positivo en la calidad del reporte de ECAs en el campo de la acupuntura^{241,242}, a pesar de existir también alguna evidencia que apunta a lo contrario^{243,244}.

A la vista de estos resultados, cabe preguntarse si la existencia de guías o extensiones adaptadas a intervenciones específicas (por ejemplo, la TM) podría mejorar la calidad del reporte e, indirectamente, también los métodos de investigación. De hecho, ya existen guías específicas para el reporte de estudios con terapia de manipulación espinal: *Consensus on Interventions Reporting Criteria List for Spinal Manipulative Therapy* (CIRCLe SMT)²⁴⁵ y con ejercicio terapéutico: *The Consensus on Exercise Reporting Template* (CERT)²⁴⁶. Esta última ha demostrado obtener un buen acuerdo entre revisores^{247,248} y existe una demanda generalizada en este campo para que se extienda su uso^{246,249–251}. Una supuesta extensión específica para el campo de la TM debería contemplar las particularidades de estos abordajes. A este respecto, hay que reconocer que al incorporar en la última versión de la extensión CONSORTnpt¹²² gran parte de los ítems propuestos por la guía TIDieR¹⁹⁰, esta extensión es ahora mucho más apropiada para su uso en el campo de la TM. Además, en el caso de considerar la creación de una nueva extensión, debe tenerse en cuenta la reducción de la carga de trabajo potencial para los autores a favor de una mayor adherencia¹¹⁹.

5.4 Fortalezas, limitaciones e implicaciones

5.4.1 Fortalezas

La principal fortaleza de este trabajo de tesis hace relación a la exhaustiva extracción de datos realizada a una extensa muestra de ECAs en el campo de la TM. A pesar de la existencia de algún trabajo similar en la literatura científica¹⁰⁰, esta tesis constituye actualmente el análisis metodológico más extenso (tanto por el periodo temporal cubierto como por el volumen de información extraída) que jamás se ha realizado en el campo de las TMs. Cabe destacar también el análisis de aspectos novedosos muy poco estudiados hasta la fecha cuyos resultados han propiciado consideraciones y puntos de vista distintos poco corrientes en los análisis metodológicos. Tanto el análisis de las limitaciones auto-reportadas (realizado en el segundo trabajo) como el análisis del reporte gráfico de los resultados (realizado en el tercer trabajo) constituyen buenos ejemplos de ello. La participación de un equipo revisor multidisciplinar constituido tanto por expertos en el campo clínico específico como por expertos metodólogos constituye otra fortaleza de este trabajo pues aporta fiabilidad a los resultados obtenidos y la adecuación del análisis al campo de las TMs^{252,253}. Cabe también mencionar que los artículos que conforman este compendio, más el artículo incluido en el anexo, han sido todos ellos publicados en revistas indexadas en el ámbito internacional, siendo especialmente destacables los dos primeros artículos publicados en la prestigiosa revista *Journal of Clinical Epidemiology* (Q1 / ISSN: 0895-4356 / IF (2020): 6.43). Sin embargo, a juicio del autor de esta tesis doctoral, la fortaleza más importante de esta tesis doctoral reside en el hecho de que constituye un punto de partida sobre el que trabajar en la mejora de la calidad de la evidencia en TM.

5.4.2 Limitaciones

Los artículos incluidos en este compendio están afectados por distintas limitaciones que constituyen, a su vez, las potenciales debilidades de esta tesis doctoral. La primera potencial limitación apunta a la representatividad de la muestra de artículos seleccionada y a la validez del método de muestreo empleado. El muestreo aleatorio simple fue el método escogido para la selección de artículos y se aseguró la potencia estadística necesaria a partir de un cálculo formal previo del tamaño de la muestra. Otras medidas para asegurar la representatividad de la muestra fueron el uso de una estrategia de búsqueda validada para identificar estudios en TM¹³², realizar la búsqueda en la base de datos más apropiada para identificar ECAs en el campo de la fisioterapia y las TMs¹²⁹⁻¹³¹ y realizar todo el proceso bajo la supervisión de un documentalista experto en la realización de revisiones sistemáticas. Los resultados de la búsqueda se aleatorizaron y se seleccionaron ordenadamente las primeras referencias que cumplían los criterios de inclusión. Si bien esta estrategia presenta pocas fisuras en relación al primer trabajo donde se comparaban dos espacios temporales, podría considerarse algo menos robusta para su utilización posterior en el segundo y tercer trabajo realizados con una muestra ampliada. Otra posible limitación asociada a la muestra fue que la selección de artículos no fue restringida a aquellas revistas manifiestamente adherentes a la declaración CONSORT. Sin embargo, nuestros resultados no difieren del único trabajo similar realizado hasta la fecha analizando artículos publicados en revistas adherentes a CONSORT¹⁰⁰. Otras potenciales limitaciones de los trabajos incluidos refieren a los formularios de extracción de datos utilizados. Respecto al primer trabajo, si bien los elementos incluidos en las guías de reporte estandarizadas (especialmente CONSORTnpt y TIDieR) formaron la matriz principal de la extracción de datos, estas guías no se utilizaron de forma sistemática y el doctorando junto con sus directores acordaron la inclusión de un número considerable de ítems adicionales. Sin embargo, esta decisión resultó ser un acierto puesto que en el transcurso de la revisión la guía CONSORTnpt fue actualizada incorporando varios de los ítems adicionales incluidos en el formulario utilizado. Por otro lado, el análisis realizado tanto en el segundo como en el tercer trabajo se basan fundamentalmente en hojas de extracción de datos confeccionadas *ad-hoc* dada la ausencia de guías oficiales o estandarizadas a tal efecto. Para minimizar potenciales sesgos se tomaron varias iniciativas durante su confección basadas en el consenso entre el grupo revisor y en el pilotaje de los cuestionarios con relación a su comprensibilidad. Sin embargo, al no ser cuestionarios formalmente validados, podrían haber ocurrido inconsistencias en la extracción de datos introduciendo potenciales sesgos.

5.4.3 Implicaciones

Las potenciales implicaciones de esta tesis doctoral se circunscriben principalmente al ámbito investigador y académico, aunque indirectamente persiguen impactar en la práctica clínica en tanto que pretenden contribuir a la mejora de la calidad de la evidencia científica en el campo de las TMs. La primera y principal contribución de esta tesis doctoral, consiste en ofrecer un sumario de las principales deficiencias metodológicas que impactan negativamente en la calidad de la evidencia de los ECAs en TM y mostrar la limitaciones más comunes que afrontan los investigadores en este campo. Conocer y entender esta información, constituye el primer paso para la búsqueda de estrategias que permitan una mejora de la calidad de la evidencia y del reporte. La segunda contribución

de esta tesis doctoral (resumida en la tabla 3) ha consistido en recolectar y presentar las recomendaciones y sugerencias propuestas por varios autores para dar respuesta a dichas deficiencias.

Sin embargo, el análisis de nuestros resultados bajo una perspectiva más global, propicia algunas reflexiones en relación a la investigación futura en este campo. Como ha quedado demostrado, la naturaleza compleja de las TMs confronta de forma inevitable con algunos aspectos de la arquitectura metodológica del ECA y es por ello que, más allá de tratar de superar las deficiencias identificadas, también es preciso contar con adaptaciones y alternativas al ECA si se pretende evaluar de forma efectiva y completa una intervención manual. Las pruebas recolectadas en la realización de esta tesis doctoral nos invitan a dibujar el futuro investigador de las TMs en 3 direcciones complementarias: i) aumentar el rigor metodológico y hacer uso de las recomendaciones existentes para mejorar la calidad y el reporte de los ECAs de corte más explicativo, ii) fomentar el uso de variantes y adaptaciones al ECA tradicional que ofrezcan un mejor encaje entre los requerimientos metodológicos y la naturaleza de las TMs (ej: ECAs pragmáticos¹⁸¹, ECAs de preferencia en 2 etapas²⁵⁴) y iii) establecer y hacer uso de diseños y marcos investigadores alternativos al ECA que permitan avanzar en el estudio de los mecanismos y efectos de la TM^{49,95}.

Finalmente y, como aportación adicional, el uso exhaustivo que se ha hecho en este trabajo de las guías de reporte ha permitido analizar su utilidad y contribución a una mejor literatura científica. Si, de forma general, consideramos su utilidad incuestionable, es importante remarcar el poco impacto que han tenido en algunos campos particulares como el de las TMs. Si bien la principal responsabilidad en la no mejora del reporte recae, sin duda, en el propio colectivo de los investigadores, los resultados obtenidos invitan a reflexionar sobre la necesidad de mejorar y/o adaptar estas herramientas a cada campo particular.

6. CONCLUSIONES

6. Conclusiones

1. A pesar del incremento de la producción científica observada en el campo de las terapias manuales en las últimas décadas, la calidad metodológica de los ensayos clínicos aleatorizados sigue siendo baja.
2. Los ensayos clínicos aleatorizados que evalúan intervenciones manuales son reportados de forma deficiente incluso años después de la disponibilidad de guías estandarizadas y la adopción de estas por parte de la mayoría de revistas especializadas en este campo.
3. El reporte de los ensayos clínicos aleatorizados en terapia manual es generalmente incompleto, lo que compromete de forma importante su interpretación, la evaluación de su calidad y su potencial aplicabilidad. Las principales deficiencias identificadas se relacionan con los siguientes elementos metodológicos: i) el tamaño de las muestras, ii) las intervenciones control, iii) las intervenciones experimentales, iv) las estrategias de enmascaramiento, v) la duración del estudio y su seguimiento y vi) las medidas de resultado.
4. El análisis de las limitaciones autoreportadas por parte de los autores de los ensayos clínicos aleatorizados en terapia manual confirma la existencia de dificultades metodológicas relacionadas especialmente con el tamaño de las muestras, la duración de los estudios y su seguimiento, y el uso de controles inadecuados.
5. El reporte de los resultados obtenidos en los ensayos clínicos aleatorizados en terapia manual es generalmente deficiente así como su presentación gráfica que a menudo es incompleta y carece de claridad visual.
6. Se constata que los déficits metodológicos en los ensayos clínicos aleatorizados en terapia manual, aunque se han producido algunas mejoras desde la introducción de guías estandarizadas para el reporte, siguen estando muy presentes. Ello pone de relieve las dificultades que entraña el diseño y realización de estudios experimentales de alta calidad para evaluar la eficacia de una intervención compleja no-farmacológica como la terapia manual.
7. En el campo de la terapia manuales, es necesario seguir promoviendo un reporte completo y estandarizado de los ensayos clínicos aleatorizados, así como desarrollar y promover métodos que permitan mejorar su calidad.

7. ABREVIATURAS

7. Abreviaturas

TM: *Terapia Manual*

TMs: *Terapias Manuales*

EEUU: *Estados Unidos*

PEDro: *Physiotherapy Evidence Database*

ISI: *Institute for Scientific Information*

CINAHL: *Cumulative Index to Nursing and Allied Health Literature*

MEDLINE: *Medical Literature Analysis and Retrieval System Online*

EMBASE: *Excerpta Medica database*

MEsH: *Medical Subject Heading*

IC: *Intervención Compleja*

ICs: *Intervenciones Complejas*

MRC: *Medical Research Council*

ECA: *Ensayo Clínico Aleatorizado*

ECAs: *Ensayos Clínicos Aleatorizados*

CONSORT: *Consolidated Standards for Reporting Trials*

CONSORTnpt: *extensión CONSORT para el reporte de intervenciones no-farmacológicas*

TIDieR: *Template for Intervention Description and Replication*

LARs: *Limitaciones Auto-Reportadas*

pre-C: *Pre-CONSORTnpt*

post-C: *Post-CONSORTnpt*

CENTRAL: *Cochrane Central Register of Controlled Trials*

MCID: *Valor Mínimo de Diferencia Clínica Importante*

SD: *Standard Desviation*

STRICTA: *Standards for Reporting Interventions in Clinical Trials of Acupuncture*

CIRCLE SMTP: *Consensus on Interventions Reporting Criteria List for Spinal Manipulative Therapy*

CERT: *Consensus on Exercise Reporting Template*

CRESS: *Centre of Research in Epidemiology and Statistics*

ITT: *Intention-To-Treat*

PRECIS: *PRagmatic Explanatory Continuum Indicator Summary*

SPIRIT: *Standard Protocol Items: Recommendations for Interventional Trials*

COMET: *Core Outcome Measures in Effectiveness Trials*

PROMs: *Patient-Reported Outcome Measures*

InsPECT: *Instrument for reporting Planned Endpoints in Clinical Trials*

8. ANEXOS

8. Anexos

8.1 Anexo 1: Estrategia de Búsqueda. (*Basada en Pillastrini et al. 2014*)

8.2 Anexo 2: Hoja de extracción de datos con instrucciones (Trabajo 1)

8.3 Anexo 3: Hoja de extracción de datos (Trabajo 3)

8.4 Anexo 4: Publicación adicional:

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;24:85-89. doi:10.1016/j.math.2016.03.004

8.1 Anexo 1: Estrategia de Búsqueda. *(Basada en Pillastrini et al. 2014)*

#1 (Chiropractic[MH] OR Manipulation, Osteopathic[MH] OR Musculoskeletal Manipulations[MH] OR Chiropractic OR Joint Mobilization* OR Manipulative OR Manual Therap* OR "Muscle Strengthening" OR "Muscle Stretching" OR Myofascial* OR Osteopathic Manipulation* OR "Proprioceptive Neuromuscular Facilitation" OR Spinal Manipulation* OR "Static Stretching" OR Trigger Point*) NOT (animals[MH] NOT humans[MH])

#2 randomized controlled trial[pt]

#3 #1 AND #2

#4 #1 AND #2 Filters: Publication date from 2000/01/01 to 2008/12/31

#5 #1 AND #2 Filters: Publication date from 2009/01/01 to 2015/01/01

8.2 Anexo2: Hoja de extracción de datos con instrucciones (Trabajo1)

Question and instructions (Internal validity and reliability)	Answer
Are the inclusion criteria well described? (There is enough information to clearly determine the study population)	YES NO
Are the criteria for the selection of therapists described? (Information about experience, academic qualifications, career ... etc. To answer YES, at least, the combination of the basic profession + professional experience or base profession + specialty is required)	YES NO
Is there a description of the center where the intervention is performed? (single-center)	YES NO
Are the selection criteria of the centers described? (multicenter)	YES NO NA
Are the specific goals or hypotheses well described? (Is there a section or paragraph specifically aimed at describing the objective of the study?)	YES NO
Are primary outcome measures described? (They must be explicitly indicated and named as "primary" including their description, how and when they were obtained) * If this information is only showed as a table in the results section or it is not categorized as "primary", the answer is NO	YES NO
Are the secondary outcome measures described? (Description of the measurements, how and when they were obtained. There is no need to be explicitly named as "secondary", "other variables" are accepted) * If this information is only showed as a table in the results section the answer is NO	YES NO NA
Are there any methods to improve the quality of data described? (Strategies focused on ensuring the reliability and the correct obtaining of, at least, the primary outcome)	YES NO NA
Is the information necessary to know how the size of the sample was calculated described? (at least: alpha and beta values, minimum change to be achieved, standard deviation in continuous variables)	YES NO
Is any specific method for the generation of the sequence described? (Not only informing what was done but how it was done)	YES NO
Is there any description about randomization and allocation of therapists to the study groups?	YES NO NA
Is it described how allocation concealment was assured? (If a randomization method that already ensures allocation concealment is described, the answer is YES. Common methods: centralized randomization or allocation by sealed envelopes)	YES NO
Is who generated the randomization sequence, who recruited the patients and who assigned the participants to each group described?	YES NO

Is how the masking of the participants was described?	YES NO
Is the masking of therapists described?	YES NO
Is how outcome assessor's masking was described?	YES NO
Is how the masking of those who administered co-interventions was described?	YES NO NA
Is the statistical analysis described for the primary variable?	YES NO
Are other analyses (subgroups, adjusted analysis, etc.) described? (if applicable - take into account if the analysis was planned from the beginning or not. If not, the answer is NO)	YES NO NA
Is the statistical analysis described in case of clustering by centers or therapists?	YES NO
Is a flowchart of each phase of the study included?	YES NO
Either in the diagram or in the text, was it described?: How many participants were randomly assigned? How many participants received the assigned treatment? The degree of compliance with the assigned treatment? Missings/dropouts during the study? In case of missings/dropouts, are the reasons described? How many participants were analyzed for the primary outcome? Any deviation from the planned protocol? (subjects who have received the intervention partially or totally) The number of therapists or centers by intervention? The number of patients treated by a therapist / center?	YES NO NA
Have the recruitment dates been described?	YES NO
Have the study follow-up periods been described?	YES NO
Are the demographic and group baseline information described?	YES NO
Is the number of participants analyzed in each group described?	YES NO
Is it mentioned that an "intention-to-treat" analysis has been performed?	YES NO
Regarding missings/follow-ups, are the imputation criteria described? (What was the statistical treatment of those missings/follow-ups in the analysis - assumptions)	YES NO NA
Are the estimated effect size and its precision described for the primary outcome? (The result must be presented either graphically or narratively but always including the estimation of the effect size and its precision)	YES NO
Was the reporting of any other analysis considered, including subgroup or adjusted analysis? (Differentiating pre-specified explorations)	YES NO
Were the adverse effects considered? (If there is no mention, the answer is NO)	YES NO

Question and instructions (Intervention description)	Answer	
Provide the name or a phrase that describes the intervention	YES	NO
Describe any rationale, theory, or goal of the elements essential to the intervention	YES	NO
Materials: Describe any physical or informational materials used in the intervention, including those provided to participants or used in intervention delivery or in training of intervention providers. Provide information on where the materials can be accessed (such as online appendix, URL)	YES	NO
Procedures: Describe each of the procedures, activities, and/or processes used in the intervention, including any enabling or support activities	YES	NO
For each category of intervention provider (such as psychologist, nursing assistant), describe their expertise, background, and any specific training given	YES	NO
Describe the modes of delivery (such as face to face or by some other mechanism, such as internet or telephone) of the intervention and whether it was provided individually or in a group	YES	NO
Describe the type(s) of location(s) where the intervention occurred, including any necessary infrastructure or relevant features	YES	NO
Describe the number of times the intervention was delivered and over what period of time including the number of sessions, their schedule, and their duration, intensity, or dose	YES	NO
If the intervention was planned to be personalized, titrated or adapted, then describe what, why, when, and how	YES	NO
If the intervention was modified during the course of the study, describe the changes (what, why, when, and how)	YES	NO
Planned: If intervention adherence or fidelity was assessed, describe how and by whom, and if any strategies were used to maintain or improve fidelity, describe them	YES	NO
Actual: If intervention adherence or fidelity was assessed, describe the extent to which the intervention was delivered as planned	YES	NO

8.3 Anexo3: Hoja de extracción de datos (Trabajo3)

Is the study hypothesis specified? <i>Yes, explicitly</i> <i>No, but it is implicit</i> <i>Not specified (cannot be determined)</i>
If the hypothesis can be determined, it is: <i>Superiority</i> <i>Non-inferiority</i>
Is a primary variable described? <i>Yes (explicitly specified)</i> <i>No (but it can be identified from the sample size calculation)</i> <i>No</i>
Variable type (n = 100) <i>Continuous</i> <i>Dichotomous</i> <i>Ordinal category</i>
Follow-up (n = 100) <i>Short term (≤ 3 months)</i> <i>Long term (> 3 months)</i>
Statistical analysis (n = 100) <i>ANOVA</i> <i>Mixed repeated measures model</i> <i>Mean difference</i> <i>T test</i> <i>Chi-Square</i> <i>Risk Ratio, Odds Ratio or Hazard Ratio</i> <i>ANCOVA</i> <i>Mann-Whitney U test</i> <i>Regression models</i> <i>Other</i>
Was the result for the main variable presented graphically? <i>Yes</i> <i>No</i>
Was a sample size calculation performed? <i>Yes</i> <i>No</i>
What information is given to evaluate how this calculation was made? <i>Alpha value</i> <i>Beta value</i> <i>Minimal change to be achieved with the intervention</i> <i>Standard deviation in continuous variables</i> <i>Expected rate or mean in the control group</i> <i>Delta margin (for non-inferiority studies)</i>
Is there sufficient information to determine whether there was a loss of subjects during the study? <i>Yes</i> <i>No</i>

In the event that some subjects were lost to follow-up, did they appear quantified (n°) in the article? Yes No Not applicable (no losses)
If some subjects were lost to follow-up, were reasons described? Yes No Not applicable (no losses)
If some subjects were lost to follow up, were the criteria for imputation of these missing data described? Yes, analysis by protocol Yes, other reasons No Not applicable (no losses)
Is the measure of the intra-group effect reported? Measure of effect is reported Only data from each evaluation is reported
Is a Comparative Effect Measurement reported? No effect or precision measure (p-value) No measure of effect, but p-value Effect measure without p-value Measure of effect and p-value
According to the authors, were the results favourable? (significant difference) Yes No
Regarding the results, is a value of clinical relevance specified? Yes, explicitly Yes, a magnitude of change is mentioned in the sample calculation No
Based on the previous question: was the result clinically relevant? Yes No A threshold of clinical relevance is not specified
Regarding completeness of graphics. Are the following characteristics met? (Yes or no) Number of subjects is discernible for each graph element Title (explicitly detailed) x axis, y axis titles (explicitly detailed) x axis, y axis labels (i.e., for tick marks or categories, label and units clear) Variance meaning defined (e.g., standard error or 95% CI) Self-explanatory (all defined data elements, including legend)
Regarding the visual clarity of graphics, are the following characteristics absent (Yes or no) Numerical distortion (scale problems or improperly scaled axes or improperly ranged axes) Chart junk (cross-hatching patterns or dark/thick/unnecessary grid lines or text labels in nonhorizontal orientation) Readability issues (error bars too cluttered or superimposition of data elements or display too small to see symbols or numeric/textual redundancy) Others (improperly connected points, labels too small to read or unclear to which graph item label refers).

8.4 Anexo4: Publicación adicional



Professional issue

Using the template for intervention description and replication (TIDieR) as a tool for improving the design and reporting of manual therapy interventions

Gerard Alvarez^{a, b, 1, *}, Francesco Cerritelli^a, Gerard Urrutia^{b, c}^a Clinical-based Human Research Department, Centre for Osteopathic Medicine – C.O.M.E. Collaboration, Pescara, Italy^b Iberoamerican Cochrane Centre – Biomedical Research Institute Sant Pau (IIB Sant Pau), Barcelona, Spain^c CIBEResp, Barcelona, Spain

ARTICLE INFO

Article history:

Received 12 December 2015

Received in revised form

1 March 2016

Accepted 3 March 2016

Keywords:

Research report

Checklist

Reproducibility of results

Musculoskeletal manipulations

Manual therapy

ABSTRACT

The detailed reporting of any research intervention is crucial to evaluate its applicability into a routinely practice-based context. However, it has been estimated that, especially in non-pharmacological interventions, the published literature typically includes incomplete intervention details. In the field of manual medicine, where interventions are delivered with a high degree of individualization and variability, poorly reported studies could compromise internal and external validity of the results. Among the various initiatives that have been undertaken to improve the intervention description, the **Template for Intervention Description and Replication (TIDieR)** has to be highlighted as the most promising. **TIDieR** offers both to researchers and clinicians a helpful and comprehensive guidance on how manual therapy interventions have to be designed and reported, taking into account the clinical complexity of manual therapy and the need to satisfy research gold standards.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

The detailed reporting of any research intervention in clinical trials is crucial to evaluate the applicability of the findings into a routinely practice-based context (external validity). Over the past decade, biased under-reporting and over-reporting of research has increasingly been acknowledged as unacceptable on both scientific and ethical grounds. However, new research is useless in many cases because of inadequate attention to important elements of study design. As a matter of fact, if clinicians are to be expected to implement treatments that have been shown in research to be useful, they need adequate descriptions of the interventions assessed. Without this information, clinical research loses its potential utility in improving patient care and involves a waste of resources (Chalmers and Glasziou 2009; Glasziou et al., 2014).

It has been estimated that, in any medical field, as much as 60% of the published literature reported incomplete intervention

details (Glasziou et al., 2008). A recent study that assessed 98 published clinical trials with public funding from the UK (Douet et al., 2014), showed that details of key components of the intervention were missing in 69.4% of cases. This problem, common for all types of treatments and interventions, is significantly worse for non-pharmacological trials where, as few as the 29–39% of interventions were described adequately compared to the 67% of drug interventions (Glasziou et al., 2008; Hoffmann et al., 2013). Other studies, although using different criteria to assess the “usability” of the description of interventions, have also shown the same shortcomings (Glasziou et al., 2008; Schroter et al., 2012). This will generally raise questions on how studies could be reproducible as well as how the external and internal validity of outcomes can be obtained. The generalizability of results would be significantly affected by intrinsic reporting biases. Moreover, the likelihood of translating the scientific results, with such biases, into the clinical practise is currently significantly impaired thus affecting the impact of therapies on the “health market” and consequently on “treatment choice”. Several authors have recently made the research community aware of such inconsistencies and proposed a number of recommendations for improvement, which include the change of the current research system to encourage better and more

* Corresponding author. Iberoamerican Cochrane Center, Hospital de la Santa Creu i Sant Pau, C/ Sant Antoni Maria Claret, 167, Pavelló 18, planta 0, 08025 Barcelona, Spain.

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

¹ Gerard Alvarez is a PhD student on Biomedical Research Methodology and Public Health in the Medical Department of the Universitat Autònoma de Barcelona.

complete reporting (Glasziou et al., 2014), focussing in particular on interventions (Hoffmann et al., 2014).

In fact, over the last 20 years, a set of general guidelines have been developed to promote a better and more consistent reporting of research. Between 1996 and 2010, the Consolidated Standards for Reporting Trials (CONSORT) statement and its revised versions were published to significantly improve the quality of clinical trials reporting in scientific journals (Plint et al., 2006). Indeed, focussing on intervention details, the CONSORT Statement evolved from the inclusion of one general item (CONSORT 2001 revised), to a larger and more detailed item in the 2008 extension of CONSORT for Trials Assessing of Non-pharmacologic Treatments (Boutron et al., 2008) and the 2010 guidelines (Moher et al., 2010). Notwithstanding this progress, specific recommendations on how to report interventions remained very limited (see Fig. 1).

A step closer to an appropriate reporting form was made by the Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) (Chan et al., 2013). This is a more recent guideline for protocols (see Table 1), which was released to specifically improve the design of clinical trials. Within the context of the 33-items SPIRIT checklist, Chan et al. (2013) dedicated a single multi-composite item (item 11) to the description of the intervention. This should be considered a consistent step forward towards an adequate and robust description, in the light of the growing awareness regarding intervention-reporting guidelines. However, it should be highlighted again, that the details included seemed to be insufficient to adequately describe interventions, especially for non-pharmacological treatments.

Among the various initiatives that have been undertaken to improve the intervention description, the Template for Intervention Description and Replication (TIDieR) (Hoffmann et al., 2014) has to be highlighted as the most promising. The checklist is an extension of the CONSORT 2010 (item 5) statement (Moher et al., 2001) and SPIRIT 2013 (item 11) (Chan et al., 2013) and focuses specifically on interventions. TIDieR consists of a checklist with 12 items and specific guidelines developed by an international group of experts and stakeholders. Those items represent the minimum information recommended for describing both (co-)interventions and comparisons (see Table 2). Beyond those 12 items, any additional information that can improve intervention replicability has to be included (if possible in the primary paper, if not as supplementary material). Checklist outlines the procedures to systematically report the rationale (item “Why”) behind the use of the intervention and the materials and procedures planned (items “What”, “How”, “Where”, “When and How Much”, “Tailoring”, “Modifications and How Well”). The primary target is to improve treatment reporting in clinical trials, however, TIDieR could be considered a substantial support for describing any type of intervention within any type of study design. This is relevant for non-

pharmacological trials but could be considered essential for manual therapy research.

2. Relevance of TIDieR for manual medicine

Considering the state of the art of manual medicine and manual therapies (MTs), scientific literature has been increasing significantly during the last few decades. Research findings seemed to demonstrate the effectiveness of several types of manual therapies in different clinical fields (Alcantara et al., 2011; Dobson et al., 2012; Pennick and Liddle, 2013; Cicchitti et al., 2015). However, considering the evidence-based health practice, there still are concerns regarding the following issues: appropriateness of using MTs in the context of complex interventions (Dobson et al., 2012; Pennick and Liddle, 2013); safety of procedures (Gouveia et al., 2009; Carnes et al., 2010; Hunsinger et al., 2014; Cicchitti et al., 2015); cost-effectiveness and cost-utility (Canter et al., 2005; Tsertsvadze et al., 2014; Cerritelli et al., 2015); and the consequent inclusion into national health care systems (Canter et al., 2005).

On the one hand, the political and lobbyist local scenario could influence the evolution of health systems and the effectiveness of multidisciplinary collaboration, although cross-disciplinary partnerships are considered a key part of science nowadays (Knapp et al., 2015). On the other hand, scientists reported that the quality of research is debatable (Jäkel and Hauenschild, 2012; Franke et al., 2015), reflecting the insufficient quality of reporting in the general scientific literature (Glasziou et al., 2014).

In the clinical context, MTs are session-based treatment plans where dosage (i.e., frequency, intensity) can vary significantly. In fact, in patients with low back pain, the variance of sessions seemed to be more associated to demographic factors and patients' complaint attitudes compared to type of treatment and therapists peculiarities (Swinkels et al., 2005). Concerning the correlation between dosage and pain, some authors reported no differences in pain pressure thresholds, despite the rate or amplitude of manoeuvre (Krouwel et al., 2010; Willett et al., 2010). Conversely, other authors pointed out that specific sets of mobilizations produced a significant reduction in pain (Pentelka et al., 2012). In addition, specific sets applied during mobilizations in chronic neck pain patients seemed to be necessary for reducing stiffness and potentially pain (Snodgrass et al., 2014) although the different type of manual manoeuvres were demonstrated to be not associated with long-term outcome effects (Izquierdo Perez et al., 2014). Conversely, the number of sessions in MTs treatments can have relevant cost-effective implications (Licciardone, 2014). Due to the degree of individualization and the high variability in the procedures (Snodgrass et al., 2006, 2007; Gorgos et al., 2014), both the description of each manoeuvre and how it is applied are mandatory in MTs studies to warrant applicability, replicability and health-care benefit assessment. TIDieR covers all these aspects including in the checklist item 4 (“What – procedures”) and item 8 (“When and How Much”), which address the full and detailed description of intervention in terms of explanation of the procedure and dosage.

Another key aspect in MTs and other non-pharmacological interventions is the therapist profile. Notwithstanding the usual cognitive, affective and psychomotor practitioners' abilities (Sizer et al., 2007, 2008), it has been suggested that musculoskeletal therapists should develop additional skills including research awareness, critical appraisal or educational capacities to obtain better clinical outcomes (Moore and Jull, 2002). The description of the intervention provider and their competencies appear, therefore, to be fundamental. TIDieR addresses this aspect in item 5 (“Who”) where the background, expertise and any training given to therapists are required.

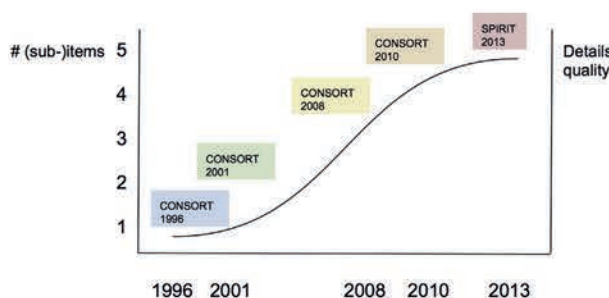


Fig. 1. Intervention description details: evolution of the different versions of CONSORT statement and SPIRIT.

Table 1
Guidelines highlighting the differences on reporting the intervention descriptions.

CONSORT statement 1996	Item: “Describe planned interventions and their timing.”
CONSORT Statement (revised) 2001	Item 4: “Precise details of the interventions intended for each group and how and when they were actually administered.”
Extension of CONSORT for Trials Assessing Nonpharmacologic Treatments 2008	Item 4: “Precise details of both the experimental treatment and comparator. The item was disaggregated into 3 subsections: item 4A) Description of the different components of the interventions and, when applicable, descriptions of the procedure for tailoring the interventions to individual participants; item 4B) Details of how the interventions were standardized; and item 4C) Details of how adherence of care providers with the protocol was assessed or enhanced.”
CONSORT Statement (revised) 2010	Item 5: “The interventions for each group with sufficient details to allow replication, including how and when they were actually administered”
CONSORT 2010 explanation & elaboration paper	“As for the description of the interventions, this document stresses the need to describe the interventions for each group with sufficient details to allow replication, including how and when they were actually administered. More particularly, it is also mentioned the need to describe in detail the control intervention, the characteristics of the placebo and the way it was prepared, the “usual treatment” that is provided as a control group and when combined interventions are used. It is also recommended in some specific cases a description of who implements the intervention as the therapist itself may be part of the intervention (eg, surgery). Finally, the paper recalls that authors should give details of when the interventions are applied (timing) and the duration, especially when it consists of multiple components.”
SPIRIT 2013	Item 11: “Interventions: 11a) Interventions for each group with sufficient detail to allow replication, including how and when they will be administered; 11b) Criteria for discontinuing or modifying allocated interventions for a given trial participant (eg, drug dose change in response to harms, participant request, or improving/worsening disease); 11c) Strategies to improve adherence to intervention protocols, and any procedures for monitoring adherence (eg, drug tablet return, laboratory tests); 11d) Relevant concomitant care and interventions that are permitted or prohibited during the trial.”

The interest in MTs efficacy (specific treatments effects) on targeted subgroups of patients has been increasing during recent years (Fersum et al., 2010; Kent et al., 2010; Slater et al., 2012). From a research perspective, some authors argued that subgrouping by patient profile and tailoring interventions lead to better outcomes (Fersum et al., 2010; Schafer et al., 2011, 2012), whereas other scientists showed inconclusive data (Hoeksma et al., 2005; Kent et al., 2005, 2010). However, in the clinical practice, MTs are mainly applied with a high degree of tailoring. This debating aspect is covered on item 9 (“Tailoring”) where the adaptation, modification and tailoring of intervention according to patient's profile is taken into account. Interestingly, including this item, the **TIDieR** guide makes a significant advance compared to other guidelines, recognizing the peculiarities of some interventions — i.e., person-based

or person-centred treatment including MTs and other non-pharmacological interventions- and claiming for a detailed treatment description. While experimental designs such as randomized clinical trials (RCTs) have been criticised by some MTs researchers (Littlewood, 2011; Milanese, 2011) as a type of research methodology not applicable in a clinical-based context, **TIDieR** can be seen as a useful tool to find a better balance between internal and external validity of MTs studies.

Finally, adherence and compliance to protocol are other key aspects for a comprehensive study. Non-pharmacological interventions such as MTs are commonly used for some chronic conditions (Reimold and Chandran, 2014; Chang et al., 2015) whose treatment can last several weeks. However, the success of interventions or research studies, which involves long-term

Table 2
Items included in the Template for Intervention Description and Replication (TIDieR) checklist (Hoffmann et al., 2014).

No.	Item	Details
1	Brief name	Provide the name or a phrase that describes the intervention
2	Why	Describe any rationale, theory, or goal of the elements essential to the intervention
3	What materials	Describe any physical or informational materials used in the intervention, including those provided to participants or used in intervention delivery or in training of intervention providers. Provide information on where the materials can be accessed (such as online appendix, URL)
4	What procedures	Describe each of the procedures, activities, and/or processes used in the intervention, including any enabling or support activities
5	Who provided	For each category of intervention provider (such as psychologist, nursing assistant), describe their expertise, background, and any specific training given
6	How	Describe the modes of delivery (such as face to face or by some other mechanism, such as internet or telephone) of the intervention and whether it was provided individually or in a group
7	Where	Describe the type(s) of location(s) where the intervention occurred, including any necessary infrastructure or relevant features
8	When and how much	Describe the number of times the intervention was delivered and over what period of time including the number of sessions, their schedule, and their duration, intensity, or dose
9	Tailoring	If the intervention was planned to be personalised, titrated or adapted, then describe what, why, when, and how
10 ^a	Modifications	If the intervention was modified during the course of the study, describe the changes (what, why, when, and how)
11	How well planned	If intervention adherence or fidelity was assessed, describe how and by whom, and if any strategies were used to maintain or improve fidelity, describe them
12 ^a	How well enacted	If intervention adherence or fidelity was assessed, describe the extent to which the intervention was delivered as planned

^a If checklist is completed for a protocol, these items are not relevant to protocol and cannot be described until the study is complete.

outcomes, was shown to be influenced by the degree by which patients accurately follow the study protocol (World Health Organization (WHO), 2003; Robiner, 2005; Deutscher et al., 2009). Adherence and compliance are, therefore, important features for setting up high-quality experimental studies. Again, **TIDieR** recognises these aspects and includes 2 items (11 and 12 – “How well”) to fulfil these needs (“Please refer to Table 2 here”).

3. Implications

Improvements in the general reporting of manual interventions could have several implications. These improvements could result in a synergistic and skilful combination of different disciplines, which would help gain insights beyond current borders and generate novel solutions to complex problems. In addition, the combination of better methods applied and how they are reported can have a meaningful impact on the planning and conduction of further research in manual medicine as well as on data obtained and interpretation. As possible effects, it would lead to: a) the possibility to explore and discuss underpinning mechanisms of actions still unexplained; b) provide insights into processes that are inaccessible using current experimental techniques; c) validate predictions against experimental data, which is of utmost importance from a scientific perspective; d) suggest adaptations of health care systems to the evolution of scientific and clinical manual medicine, which can be considered an ultimate long term outcome and e) facilitate more accurate assessment in terms of cost-effectiveness and cost-utility in the implementation of these disciplines.

4. Conclusions

In general, **TIDieR** offers both researchers and clinicians a helpful and comprehensive guidance on how MT interventions have to be designed and reported taking into account the clinical complexity of MTs and the need to satisfy research gold standards. Following the **TIDieR** checklist would increase the probability of studies being replicated and reproduced, extending the external validity of results and yielding, potentially, the generalizability of benefits. Moreover, it can contribute to several other positive changes such as: 1) increased awareness of what is regarded as a complete description of intervention by authors and reviewers as well as by journals and editors; 2) better understanding of manual practice for external professionals, unfamiliar with the given manual medicines; 3) enhancing the scientific credibility of MTs; 4) increasing the likelihood of creating profitable cross-disciplinary and inter-disciplinary discussions; 5) defining clearer competencies, magnifying strengths or limitations to patients and for policymakers.

References

Alcantara J, Alcantara JD, Alcantara J. The chiropractic care of infants with colic: a systematic review of the literature. *Explore (NY)* 2011 May;7(3):168–74.

Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P, CONSORT Group. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. *Ann Intern Med* 2008 Feb 19;148(4):295–309.

Canter PH, Coon JT, Ernst E. Cost effectiveness of complementary treatments in the United Kingdom: systematic review. *BMJ* 2005 Oct 15;331(7521):880–1. British Medical Journal Publishing Group.

Carnes D, Mars TS, Mullinger B, Froud R, Underwood M. Adverse events and manual therapy: a systematic review. *Man Ther* 2010 Aug;15(4):355–63.

Cerritelli F, Pizzolorusso G, Renzetti C, Cozzolino V, D'Orazio M, Lupacchini M, et al. A multicenter, randomized, controlled trial of osteopathic manipulative treatment on preterms. In: Rubens C, editor. *PLoS One*, 10(5). Public Library of Science; 2015. e0127370.

Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. *Lancet* 2009 Jul 4;374(9683):86–9.

Chan A-W, Tetzlaff JM, Gøtzsche PC, Altman DG, Mann H, Berlin JA, et al. SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials. *BMJ* 2013 Jan 08;346(15):e7586. British Medical Journal Publishing Group.

Chang K-L, Fillingim R, Hurley RW, Schmidt S. Chronic pain management: non-pharmacological therapies for chronic pain. *FP Essent* 2015 May;432:21–6.

Cicchetti L, Martelli M, Cerritelli F. Chronic inflammatory disease and osteopathy: a systematic review. In: d'Acquisto F, editor. *PLoS One*, 10(3). Public Library of Science; 2015. e0121327.

Deutscher D, Horn SD, Dickstein R, Hart DL, Smout RJ, Gutvirtz M, et al. Associations between treatment processes, patient characteristics, and outcomes in outpatient physical therapy practice. *Arch Phys Med Rehabil* 2009 Aug;90(8):1349–63.

Dobson D, Lucassen PLBJ, Miller JJ, Vlieger AM, Prescott P, Lewith G. Manipulative therapies for infantile colic. In: Dobson D, editor. *Cochrane Database Syst Rev*, 12. Chichester, UK: John Wiley & Sons, Ltd; 2012. CD004796.

Douet L, Milne R, Anstee S, Habens F, Young A, Wright D. The completeness of intervention descriptions in published National Institute of Health Research HTA-funded trials: a cross-sectional study. *BMJ Open* 2014;4(1):e003713. British Medical Journal Publishing Group.

Fersum KV, Dankaerts W, O'Sullivan PB, Maes J, Skouen JS, Bjordal JM, et al. Integration of subclassification strategies in randomised controlled clinical trials evaluating manual therapy treatment and exercise therapy for non-specific chronic low back pain: a systematic review. *Br J Sports Med* 2010 Nov;44(14):1054–62.

Franke H, Fryer G, Ostelo RWJG, Kamper SJ. Muscle energy technique for non-specific low-back pain. In: Franke H, editor. *Cochrane Database Syst Rev*, 2. Chichester, UK: John Wiley & Sons, Ltd; 2015. CD009852.

Glasziou P, Meats E, Heneghan C, Shepperd S. What is missing from descriptions of treatment in trials and reviews? *BMJ* 2008 Jun 28;336(7659):1472–4. British Medical Journal Publishing Group.

Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, Julious S, et al. Reducing waste from incomplete or unusable reports of biomedical research. *Lancet* 2014 Jan 18;383(9913):267–76.

Gorgos KS, Wasyluk NT, Van Lunen BL, Hoch MC. Inter-clinician and intra-clinician reliability of force application during joint mobilization: a systematic review. *Man Ther* 2014 Apr;19(2):90–6.

Gouveia LO, Castanho P, Ferreira JJ. Safety of chiropractic interventions: a systematic review. *Spine* 2009 May 15;34(11):E405–13.

Hoeksma HL, Dekker J, Ronday HK, Breedveld FC, Van den Ende CHM. Manual therapy in osteoarthritis of the hip: outcome in subgroups of patients. *Rheumatology* 2005 Apr;44(4):461–4.

Hoffmann TC, Eructi C, Glasziou PP. Poor description of non-pharmacological interventions: analysis of consecutive sample of randomised trials. *BMJ* 2013 Sep 10;347(1):f3755.

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:g1687.

Hunsinger M, Smith SM, Rothstein D, McKeown A, Parkhurst M, Hertz S, et al. Adverse event reporting in nonpharmacologic, noninterventional pain clinical trials: ACTION systematic review. *Pain* 2014 Nov;155(11):2253–62.

Izquierdo Perez H, Alonso Perez JL, Gil Martinez A, La Touche R, Lerma-Lara S, Commaux Gonzalez N, et al. Is one better than another?: a randomized clinical trial of manual therapy for patients with chronic neck pain. *Man Ther* 2014 Jun;19(3):215–21.

Jäkel A, Hauenchild von P. A systematic review to evaluate the clinical benefits of craniosacral therapy. *Complement Ther Med* 2012 Dec;20(6):456–65.

Kent P, Marks D, Pearson W, Keating J. Does clinician treatment choice improve the outcomes of manual therapy for nonspecific low back pain? A metaanalysis. *J Manipulative Physiol Ther* 2005 Jun;28(5):312–22.

Kent P, Mjosund HL, Petersen DHD. Does targeting manual therapy and/or exercise improve patient outcomes in nonspecific low back pain? A systematic review. *BMC Med* 2010;8(1):22.

Knapp B, Bardenet R, Bernabeu MO, Bordas R, Bruna M, Calderhead B, et al. Ten simple rules for a successful cross-disciplinary collaboration. *PLoS Comput Biol* 2015 Apr;11(4):e1004214. Public Library of Science.

Krouwel O, Hebron C, Willett E. An investigation into the potential hypoalgesic effects of different amplitudes of Pa mobilisations on the lumbar spine as measured by pressure pain thresholds (PPT). *Man Ther* 2010 Feb;15(1):7–12.

Licciardone JC. Short-term dosing of manual therapies for chronic low back pain. *Spine J* 2014 Jun 1;14(6):1085–6.

Littlewood C. The RCT means nothing to me! *Man Ther* 2011 Dec;16(6):614–7.

Milanese S. The use of RCT's in manual therapy – are we trying to fit a round peg into a square hole? *Man Ther* 2011 Jan 8;16(4):403–5.

Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *The Lancet* 2001 Apr;357(9263):1191–4.

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 Mar 23;340(1):c869.

Moore A, Jull G. Reflections on the musculoskeletal therapists' multifaceted role and influences on treatment outcomes. *Man Ther* 2002;7(3):119–20.

Pennick V, Liddle SD. Interventions for preventing and treating pelvic and back pain in pregnancy. In: Pennick V, editor. *Cochrane Database Syst Rev*, 8. Chichester, UK: John Wiley & Sons, Ltd; 2013. CD001139.

- Pentelka L, Hebron C, Shapleski R, Goldshtein I. The effect of increasing sets (within one treatment session) and different set durations (between treatment sessions) of lumbar spine posteroanterior mobilisations on pressure pain thresholds. *Man Ther* 2012 Dec;17(6):526–30.
- Plint AC, Moher D, Morrison A, Schulz K, Altman DG, Hill C, et al. Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. *Med J Aust* 2006 Sep 4;185(5):263–7.
- Reimold AM, Chandran V. Nonpharmacologic therapies in spondyloarthritis. *Best Pract Res Clin Rheumatol* 2014 Oct;28(5):779–92.
- Robiner WN. Enhancing adherence in clinical research. *Contemp Clin Trials* 2005 Feb;26(1):59–77.
- Schafer A, Hall T, Muller G, Briffa K. Outcomes differ between subgroups of patients with low back and leg pain following neural manual therapy: a prospective cohort study. *Eur Spine J* 2011 Mar;20(3):482–90.
- Schroter S, Glasziou P, Heneghan C. Quality of descriptions of treatments: a review of published randomised controlled trials. *BMJ Open* 2012 Jan 1;2(6):e001978. British Medical Journal Publishing Group.
- Sizer P, Felstehausen V, Sawyer S, Dornier L, Matthews P, Cook C. Eight critical skill sets required for manual therapy competency: a Delphi study and factor analysis of physical therapy educators of manual therapy. *J Allied Health* 2007;36(1):30–40.
- Sizer P, Sawyer S, Felstehausen V, Couch S, Dornier L, Cook C. Intrinsic and extrinsic factors important to manual therapy competency development: a Delphi investigation. *J Man Manipulative Ther* 2008;16(1):e9–19.
- Slater SL, Ford JJ, Richards MC, Taylor NF, Surkitt LD, Hahne AJ. The effectiveness of sub-group specific manual therapy for low back pain: a systematic review. *Man Ther* 2012 Jun;17(3):201–12.
- Snodgrass SJ, Rivett DA, Robertson VJ. Manual forces applied during posterior-to-anterior spinal mobilization: a review of the evidence. *J Manipulative Physiol Ther* 2006 May;29(4):316–29.
- Snodgrass SJ, Rivett DA, Robertson VJ. Manual forces applied during cervical mobilization. *J Manipulative Physiol Ther* 2007 Jan;30(1):17–25.
- Snodgrass SJ, Rivett DA, Sterling M, Vicenzino B. Dose optimization for spinal treatment effectiveness: a randomized controlled trial investigating the effects of high and low mobilization forces in patients with neck pain. *J Orthop Sports Phys Ther* 2014 Mar;44(3):141–52.
- Swinkels ICS, Wimmers RH, Groenewegen PP, van den Bosch WJH, Dekker J, Van Den Ende CHM. What factors explain the number of physical therapy treatment sessions in patients referred with low back pain; a multilevel analysis. *BMC Health Serv Res* 2005;5(1):74. BioMed Central Ltd.
- Tsertsvadze A, Clar C, Court R, Clarke A, Mistry H, Sutcliffe P. Cost-effectiveness of manual therapy for the management of musculoskeletal conditions: a systematic review and narrative synthesis of evidence from randomized controlled trials. *J Manipulative Physiological Ther* 2014 Jul;37(6):343–62.
- Willett E, Hebron C, Krouwel O. The initial effects of different rates of lumbar mobilisations on pressure pain thresholds in asymptomatic subjects. *Man Ther* 2010 Apr;15(2):173–8.
- World Health Organization (WHO). Adherence to long-term therapies. 2003. p. 1–194 (accessed 12.12.15), <http://apps.who.int/iris/bitstream/10665/42682/1/9241545992.pdf>.

9. BIBLIOGRAFÍA

9. Bibliografía

1. Paris S V. A history of manipulative therapy through the ages and up to the current controversy in the United States. *J Man Manip Ther.* 2000;8(2):66-77. doi:10.1179/106698100790819555
2. Smith AR. Manual therapy: The historical, current, and future role in the treatment of pain. *ScientificWorldJournal.* 2007;7:109-120. doi:10.1100/tsw.2007.14
3. French HP, Brennan A, White B, Cusack T. Manual therapy for osteoarthritis of the hip or knee - A systematic review. *Man Ther.* 2011;16(2):109-117. doi:10.1016/j.math.2010.10.011
4. Pettman E. A history of manipulative therapy. *J Man Manip Ther.* 2007;15(3):165-174. doi:10.1179/106698107790819873
5. Lewis JR. *A.T. Still: From the Dry Bone to the Living Man.* Dry Bone Press; 2012.
6. Showalter CR. In memoriam: farewell Geoff Maitland. *J Man Manip Ther.* 2010;18(1):56-57. doi:10.1179/106698110x12640740712978
7. Chaitow L. Gregory Peter Grieve 1918–2001. *J Bodyw Mov Ther.* 2002;6(3):199. doi:10.1054/jbmt.2002.0296
8. Zusman M. The Modernisation of Manipulative Therapy. *Int J Clin Med.* 2011;02(05):644-649.
9. Oostendorp RAB. Credibility of manual therapy is at stake 'Where do we go from here?' *J Man Manip Ther.* 2018;26(4):189-192. doi:10.1080/10669817.2018.1472948
10. Lederman E. The fall of the postural-structural-biomechanical model in manual and physical therapies: Exemplified by lower back pain. *J Bodyw Mov Ther.* 2011;15(2):131-138. doi:10.1016/j.jbmt.2011.01.011
11. Dunk NM, Chung YY, Compton DS, Callaghan JP. The reliability of quantifying upright standing postures as a baseline diagnostic clinical tool. *J Manipulative Physiol Ther.* 2004;27(2):91-96. doi:10.1016/j.jmpt.2003.12.003
12. Seffinger MA, Najm WI, Mishra SI, et al. Reliability of Spinal Palpation for Diagnosis of Back and Neck Pain. *Spine (Phila Pa 1976).* 2004;29(19):E413-E425. doi:10.1097/01.brs.0000141178.98157.8e
13. Smith D. Reflecting on new models for osteopathy – it's time for change. *Int J Osteopath Med.* 2019;31:15-20. doi:10.1016/j.ijosm.2018.10.001
14. Van Trijffel E, Anderegg Q, Bossuyt PMM, Lucas C. Inter-examiner reliability of passive assessment of intervertebral motion in the cervical and lumbar spine: A systematic review. *Man Ther.* 2005;10(4):256-269. doi:10.1016/j.math.2005.04.008
15. Franke H. Re: Is a postural-structural-biomechanical model, within manual therapy, viable? A JBMT debate. *J Bodyw Mov Ther.* 2011;15(3):259-261. doi:10.1016/j.jbmt.2011.03.002
16. Wielechowski A, Feldman K. Letter to the editor: the evolution of manual therapy education: what are we waiting for? *J Man Manip Ther.* 2020;28(1):1-3. doi:10.1080/10669817.2020.1780083
17. Bialosky JE, Bishop MD, Price DD, Robinson ME, George SZ. The mechanisms of manual therapy in the treatment of musculoskeletal pain: A comprehensive model. *Man Ther.* 2009. doi:10.1016/j.math.2008.09.001

18. Bialosky JE, Beneciuk JM, Bishop MD, et al. Unraveling the Mechanisms of Manual Therapy: Modeling an Approach. *J Orthop Sports Phys Ther.* 2018;48(1):8-18. doi:10.2519/jospt.2018.7476
19. D'Alessandro G, Cerritelli F, Cortelli P. Sensitization and Interoception as Key Neurological Concepts in Osteopathy and Other Manual Medicines. *Front Neurosci.* 2016;10(503):2812-2879.
20. Collins CK, Masaracchio M, Brismée JM. The future of orthopedic manual therapy: what are we missing? *J Man Manip Ther.* 2017;25(4):169-171. doi:10.1080/10669817.2017.1358249
21. 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(5):328-337. doi:10.1016/j.joim.2019.05.004
22. Voogt L, de Vries J, Meeus M, Struyf F, Meuffels D, Nijs J. Analgesic effects of manual therapy in patients with musculoskeletal pain: A systematic review. *Man Ther.* 2015;20(2):250-256. doi:10.1016/j.math.2014.09.001
23. Pelletier R, Bourbonnais D, Higgins J. Nociception, pain, neuroplasticity and the practice of Osteopathic Manipulative Medicine. *Int J Osteopath Med.* 2018;27:34-44. doi:10.1016/j.ijosm.2017.08.001
24. Shacklock MO. Central pain mechanisms: A new horizon in manual therapy. *Aust J Physiother.* 1999;45(2):83-92. doi:10.1016/S0004-9514(14)60340-8
25. Schmid A, Brunner F, Wright A, Bachmann LM. Paradigm shift in manual therapy? Evidence for a central nervous system component in the response to passive cervical joint mobilisation. *Man Ther.* 2008;13(5):387-396.
26. Dunn SL, Olmedo ML. Mechanotransduction: Relevance to physical therapist practice— Understanding our ability to affect genetic expression through mechanical forces. *Phys Ther.* 2016;96(5):712-721. doi:10.2522/ptj.20150073
27. Chaitow L. Fascial well-being: Mechanotransduction in manual and movement therapies. *J Bodyw Mov Ther.* 2018;22(2):235-236. doi:10.1016/j.jbmt.2017.11.011
28. Ng JL, Kersh ME, Kilbreath S, Knothe Tate M. Establishing the basis for mechanobiology-based physical therapy protocols to potentiate cellular healing and tissue regeneration. *Front Physiol.* 2017;8(JUN):1-13. doi:10.3389/fphys.2017.00303
29. Tozzi P. A unifying neuro-fasciagenic model of somatic dysfunction - Underlying mechanisms and treatment - Part II. *J Bodyw Mov Ther.* 2015;19(3):526-543. doi:10.1016/j.jbmt.2015.03.002
30. Bialosky JE, Bishop MD, Penza CW. Placebo Mechanisms of Manual Therapy: A Sheep in Wolf's Clothing? *J Orthop Sport Phys Ther.* 2017;47(5):301-304. doi:10.2519/jospt.2017.0604
31. Koes BW. How to evaluate manual therapy: Value and pitfalls of randomized clinical trials. *Man Ther.* 2004;9(4):183-184. doi:10.1016/j.math.2004.04.002
32. Vielitz A. Enhance Placebo, Avoid Nocebo: How Contextual Factors Affect Physiotherapy Outcomes. *Manuelletherapie.* 2017;21(01):8-9. doi:10.1055/s-0042-123422
33. Menke JM. Manual therapy researchers are misled by natural history and placebo effects. *Spine (Phila Pa 1976).* 2014;39(23):1997. doi:10.1097/BRS.0000000000000585

34. Bialosky JE, Bishop MD, George SZ, Robinson ME. Placebo response to manual therapy: Something out of nothing? *J Man Manip Ther.* 2011;19(1):11-19. doi:10.1179/2042618610Y.0000000001
35. Rossetтини 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(1):1-15. doi:10.1186/s12891-018-1943-8
36. Di Blasi Z, Harkness E, Ernst E, Georgiou A, Kleijnen J. Influence of context effects on health outcomes: a systematic review. *Lancet.* 2001;357(9258):757-762.
37. Di Blasi Z, Kleijnen J. Context effects: Powerful therapies or methodological bias? *Eval Heal Prof.* 2003;26(2):166-179. doi:10.1177/0163278703026002003
38. Acquati A, Uberti S, Aquino A, et al. Do empathic osteopaths achieve better clinical results? An observational feasibility study. *Int J Osteopath Med.* 2019;32(February):2-6. doi:10.1016/j.ijosm.2019.02.002
39. 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(14):907-909. doi:10.1136/bjsports-2014-094468
40. Cashin AG, Lee H, Bagg MK, 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;126:106-115. doi:10.1016/j.jclinepi.2020.06.025
41. Fell DW, Burnham JF, Buchanan MJ, Horchen HA, Scherr JA. Mapping the core journals of the physical therapy literature. *J Med Libr Assoc.* 2011;99(3):202-207. doi:10.3163/1536-5050.99.3.007
42. Smith DR, Rivett DA. Bibliometrics, impact factors and manual therapy: Balancing the science and the art. *Man Ther.* 2009;14(4):456-459. doi:10.1016/j.math.2008.11.004
43. Turrillas M, Sitjà-Rabert M, Pardo H, et al. Identification and description of controlled clinical trials published in Physiotherapy journals in Spain. *J Eval Clin Pract.* 2017;23(1):29-36. doi:10.1111/jep.12441
44. Beeton K. Masterclass editorial. *Man Ther.* 2008;19(3):173-174. doi:10.1016/j.math.2014.04.003
45. Bohannon RW, Roberts D. Core journals of rehabilitation: identification through index analysis. *Int J Rehabil Res.* 1991;14(4):333-336.
46. Jones DS, Podolsky SH. The history and fate of the gold standard. *Lancet.* 2015;385(9977):1502-1503. doi:10.1016/S0140-6736(15)60742-5
47. Kane RL, Wang J, Garrard J. Reporting in randomized clinical trials improved after adoption of the CONSORT statement. *J Clin Epidemiol.* 2007;60(3):241-249. doi:10.1016/j.jclinepi.2006.06.016
48. Kaptchuk TJ. The double-blind, randomized, placebo-controlled trial: Gold standard or golden calf? *J Clin Epidemiol.* 2001;54(6):541-549. doi:10.1016/S0895-4356(00)00347-4
49. 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;6:1-9. doi:10.1186/1471-2288-6-29

50. Wade DT, Smeets RJE, Verbunt JA. Research in rehabilitation medicine: Methodological challenges. *J Clin Epidemiol*. 2010;63(7):699-704. doi:10.1016/j.jclinepi.2009.07.010
51. Kerry R. Expanding our perspectives on research in musculoskeletal science and practice. *Musculoskelet Sci Pract*. October 2017. <http://linkinghub.elsevier.com/retrieve/pii/S2468781217301534>.
52. 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(3):319-321. doi:10.23736/S1973-9087.19.05811-8
53. Arienti C, Armijo-Olivo S, Minozzi S, Lazzarini SG, Patrini M, Negrini S. 60 Methodological issues in rehabilitation research: a scoping review. *BMJ Evidence-Based Med*. 2019;24(Suppl 1):A35--A35. doi:10.1136/bmjebm-2019-EBMLive.68
54. Dieppe P. Complex interventions. *Musculoskeletal Care*. 2004;2(3):180-186. doi:10.1002/msc.69
55. Milanese S. The use of RCT's in manual therapy - Are we trying to fit a round peg into a square hole? *Man Ther*. 2011;16(4):403-405. doi:10.1016/j.math.2011.02.007
56. Licciardone JC, Russo DP. Blinding protocols, treatment credibility, and expectancy: methodologic issues in clinical trials of osteopathic manipulative treatment. *J Am Osteopath Assoc*. 2006;106(8):457-463.
57. Patterson MM. Research in OMT: What is the question and do we understand it? *J Am Osteopath Assoc*. 2007;107(1):8-11. doi:10.7556/jaoa.2007.107.1.8
58. Littlewood C. The RCT means nothing to me! *Man Ther*. 2011;16(6):614-617. doi:10.1016/j.math.2011.06.006
59. Bolton JE. Research design in chiropractic. *J Manipulative Physiol Ther*. 1999;22(7):483-487. doi:10.1016/S0161-4754(99)70038-2
60. Rosner A. Fables or foibles: Inherent problems with RCTs. *J Manipulative Physiol Ther*. 2003;26(7):460-467. doi:10.1016/S0161-4754(03)00094-0
61. Cook CE, Thigpen CA. Five good reasons to be disappointed with randomized trials. *J Man Manip Ther*. 2019;27(2):63-65. doi:10.1080/10669817.2019.1589697
62. Saracutu M, Rance J, Davies H, Edwards DJ. The effects of osteopathic treatment on psychosocial factors in people with persistent pain: {A} systematic review. *Int J Osteopath Med*. 2018;27:23-33. doi:10.1016/j.ijosm.2017.10.005
63. Djulbegovic B, Cantor A, Clarke M. The importance of the preservation of the ethical principle of equipoise in the design of clinical trials: Relative impact of the methodological quality domains on the treatment effect in randomized controlled trials. *Account Res*. 2003;10(4):301-315. doi:10.1080/714906103
64. Djulbegovic B, Clarke M. Scientific and ethical issues in equivalence trials. *J Am Med Assoc*. 2001;285(9):1206-1208. doi:10.1001/jama.285.9.1206
65. Djulbegovic B, Lacevic M, Cantor A, et al. The uncertainty principle and industry-sponsored research. *Lancet*. 2000;356(9230):635-638. doi:10.1016/S0140-6736(00)02605-2
66. Chaibi A, Šaltytė Benth J, Bjørn Russell M. Validation of Placebo in a Manual Therapy Randomized Controlled Trial. *Sci Rep*. 2015;5(1):11774. doi:10.1038/srep11774

67. 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(2):0370-0380. doi:10.1371/journal.pmed.0040061
68. Hancock MJ, Maher CG, Latimer J, McAuley JH. Selecting an appropriate placebo for a trial of spinal manipulative therapy. *Aust J Physiother*. 2006;52(2):135-138.
69. Michener LA, Kardouni JR, Lopes Albers AD, Ely JM. Development of a sham comparator for thoracic spinal manipulative therapy for use with shoulder disorders. *Man Ther*. 2013;18(1):60-64. doi:10.1016/j.math.2012.07.003
70. Puhl AA, Reinhart CJ, Doan JB, Vernon H. The quality of placebos used in randomized, controlled trials of lumbar and pelvic joint thrust manipulation—a systematic review. *Spine J*. 2017;17(3):445-456. doi:10.1016/j.spinee.2016.11.003
71. Cerritelli F, Verzella M, Cicchitti L, D'Alessandro G, Vanacore N. The paradox of sham therapy and placebo effect in osteopathy A systematic review. *Med (United States)*. 2016;95(35). doi:10.1097/MD.00000000000004728
72. Ernst E. Placebo: new insights into an old enigma. *Drug Discov Today*. 2007;12(9-10):413-418. doi:10.1016/j.drudis.2007.03.007
73. Vernon HT, Triano JJ, Ross JK, Tran SK, Soave DM, Dinulos MD. Validation of a novel sham cervical manipulation procedure. *Spine J*. 2012;12(11):1021-1028. doi:10.1016/j.spinee.2012.10.009
74. Vernon H, MacAdam K, Marshall V, Pion M, Sadowska M. Validation of a sham manipulative procedure for the cervical spine for use in clinical trials. *J Manipulative Physiol Ther*. 2005;28(9):662-666. doi:10.1016/j.jmpt.2005.07.020
75. Sedgwick P, Greenwood N. Understanding the hawthorne effect. *BMJ*. 2015;351(September):1-2. doi:10.1136/bmj.h4672
76. Boutron I, Tubach F, Giraudeau B, Ravaud P. Methodological Differences in Clinical Trials Evaluating Nonpharmacological and Pharmacological Treatments of Hip and Knee Osteoarthritis. *J Am Med Assoc*. 2003;290(8):1062-1070. doi:10.1001/jama.290.8.1062
77. Hawe P, Shiell A, Riley T. Complex interventions: How “out of control” can a randomised controlled trial be? *Bmj*. 2004;328(7455):1561. doi:10.1136/bmj.328.7455.1561
78. 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. doi:10.1186/1471-2288-11-24
79. Hafliðadóttir SH, Juhl CB, Nielsen SM, et al. Placebo response and effect in randomized clinical trials: meta-research with focus on contextual effects. *Trials*. 2021;22(1):1-15. doi:10.1186/s13063-021-05454-8
80. Zou K, Wong J, Abdullah N, et al. Examination of overall treatment effect and the proportion attributable to contextual effect in osteoarthritis: Meta-Analysis of randomised controlled trials. *Ann Rheum Dis*. 2016;75(11):1964-1970. doi:10.1136/annrheumdis-2015-208387
81. Whiteside N, Sarmanova A, Chen X, et al. Proportion of contextual effects in the treatment of fibromyalgia-a meta-analysis of randomised controlled trials. *Clin Rheumatol*. 2018;37(5):1375-1382. doi:10.1007/s10067-017-3948-3

82. Testa M, Rossetini G. Enhance placebo, avoid nocebo: How contextual factors affect physiotherapy outcomes. *Man Ther.* 2016;24:65-74. doi:<https://doi.org/10.1016/j.math.2016.04.006>
83. Rossetini G, Palese A, Geri T, Fiorio M, Colloca L, Testa M. Physical therapists' perspectives on using contextual factors in clinical practice: Findings from an Italian national survey. *PLoS One.* 2018;13(11):e0208159. doi:10.1371/journal.pone.0208159
84. Verhagen A, Yu LM. Editorial: What about sample size? *Musculoskelet Sci Pract.* 2021;54:102405. doi:10.1016/j.msksp.2021.102405
85. Sedgwick P. Randomised controlled trials: The importance of sample size. *BMJ.* 2015;350(March):1-2. doi:10.1136/bmj.h1586
86. Slater SL, Ford JJ, Richards MC, Taylor NF, Surkitt LD, Hahne AJ. The effectiveness of sub-group specific manual therapy for low back pain: a systematic review. *Man Ther.* 2012;17(3):201-212. doi:10.1016/j.math.2012.01.006
87. Pearce W, Raman S, Turner A. Randomised trials in context: Practical problems and social aspects of evidence-based medicine and policy. *Trials.* 2015;16(1):1-7. doi:10.1186/s13063-015-0917-5
88. Rothwell PM. Factors That Can Affect the External Validity of Randomised Controlled Trials. *PLoS Clin Trials.* 2006;1(1):e9. doi:10.1371/journal.pctr.0010009
89. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". *Lancet.* 2005;365(9453):82-93. doi:10.1016/S0140-6736(04)17670-8
90. Cartwright N. A philosopher's view of the long road from RCTs to effectiveness. *Lancet.* 2011;377(9775):1400-1401. doi:10.1016/S0140-6736(11)60563-1
91. Medical Research Council. MRC Developing and evaluating complex interventions. Medical Research Council. www.mrc.ac.uk/complexinterventionsguidance. Published 2006.
92. Craig P, Dieppe P, Macintyre S, Mitchie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: The new Medical Research Council guidance. *Bmj.* 2008;337(7676):979-983. doi:10.1136/bmj.a1655
93. Campbell M, Fitzpatrick R, Haines A, et al. Framework for design and evaluation of complex interventions to improve health Framework for trials of complex interventions. *Br Med J.* 2000;321(7262):694-696. doi:10.1136/bmj.321.7262.694
94. 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(10):1525-1541. doi:10.2522/ptj.20100430
95. 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):1-9. doi:10.1186/s12874-019-0736-6
96. Moore GF, Audrey S, Barker M, et al. Process evaluation of complex interventions: Medical Research Council guidance. *BMJ.* 2015;350:1-7. doi:10.1136/bmj.h1258
97. Medical Research Council. *A Framework for Development and Evaluation of RCTs for Complex Interventions to Improve Health.* London; 2000. <https://mrc.ukri.org/docu->

- ments/pdf/rcts-for-complex-interventions-to-improve-health/. Accessed August 11, 2020.
98. Safran-Norton CE, Sullivan JK, Irrgang JJ, et al. A consensus-based process identifying physical therapy and exercise treatments for patients with degenerative meniscal tears and knee OA: The TeMPO physical therapy interventions and home exercise program. *BMC Musculoskelet Disord*. 2019;20(1):1-11. doi:10.1186/s12891-019-2872-x
 99. Clar C, Tsertsvadze A, Court R, Hundt G, 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(1):12. doi:10.1186/2045-709X-22-12
 100. 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(5):241-252. doi:10.1080/10669817.2015.1119372
 101. Moseley AM, Elkins MR, Janer-Duncan L, Hush JM. The quality of reports of randomized controlled trials varies between subdisciplines of physiotherapy. *Physiother Canada*. 2014;66(1):36-43. doi:10.3138/ptc.2012-68
 102. Gonzalez GZ, Moseley AM, Maher CG, Nascimento DP, Costa L i ola da CMLOL da CM, Costa L i ola da CMLOL da CM. Methodologic Quality and Statistical Reporting of Physical Therapy Randomized Controlled Trials Relevant to Musculoskeletal Conditions. *Arch Phys Med Rehabil*. 2018;99(1):129-136. doi:10.1016/j.apmr.2017.08.485
 103. Costa LOP, Maher CG, Lopes AD, Noronha MA de, Costa LCM. Transparent reporting of studies relevant to physical therapy practice. *Brazilian J Phys Ther*. 2011. doi:10.1590/s1413-35552011005000009
 104. Van Tulder MW, Suttrop M, Morton S, Bouter LM, Shekelle P. Empirical Evidence of an Association Between Internal Validity and Effect Size in Randomized Controlled Trials of Low-Back Pain. *Spine (Phila Pa 1976)*. 2009;34(16):1685-1692.
 105. Bertizzolo L, Bossuyt PM, Atal I, Ravaud P, Dechartres A. External factors may influence Cochrane reviewers when classifying the risk of bias of original reports. *J Clin Epidemiol*. 2020;123:1-8. doi:10.1016/j.jclinepi.2020.03.008
 106. Moseley AM, Rahman P, Wells GA, et al. Agreement between the Cochrane risk of bias tool and Physiotherapy Evidence Database (PEDro) scale: A meta-epidemiological study of randomized controlled trials of physical therapy interventions. *PLoS One*. 2019;14(9):1-16. doi:10.1371/journal.pone.0222770
 107. Altman DG. Better reporting of randomised controlled trials: The CONSORT statement. *Br Med J*. 1996;313(7057):570-571. doi:10.1136/bmj.313.7057.570
 108. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340. doi:10.1136/bmj.c869
 109. Chan L, Heinemann AW, Roberts J. Elevating the quality of disability and rehabilitation research: Mandatory use of the reporting guidelines. *Am J Occup Ther*. 2014;68(2):127-129. doi:10.5014/ajot.2014.682004
 110. Moseley AM, Herbert RD, Maher CG, Sherrington C, Elkins MR. Reported quality of randomized controlled trials of physiotherapy interventions has improved over time. *J Clin Epidemiol*. 2011;64(6):594-601. doi:10.1016/j.jclinepi.2010.08.009

111. Plint AC, Moher D, Morrison A, et al. Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. *Med J Aust.* 2006;185(5):263-267. doi:pli11098_fm [pii]
112. Reveiz L, Chapman E, Asial S, Munoz S, Bonfill X, Alonso-Coello P. Risk of bias of randomized trials over time. *J Clin Epidemiol.* 2015;68(9):1036-1045. doi:10.1016/j.jclinepi.2014.06.001
113. Dechartres A, Charles P, Hopewell S, Ravaud P, Altman DG. Reviews assessing the quality or the reporting of randomized controlled trials are increasing over time but raised questions about how quality is assessed. *J Clin Epidemiol.* 2011;64(2):136-144. doi:10.1016/j.jclinepi.2010.04.015
114. Gray R, Sullivan M, Altman DG, Gordon-Weeks AN. Adherence of trials of operative intervention to the CONSORT statement extension for nonpharmacological treatments: A comparative before and after study. *Ann R Coll Surg Engl.* 2012;94(6):388-394. doi:10.1308/003588412X13171221592339
115. Nagendran M, Harding D, Teo W, et al. Poor adherence of randomised trials in surgery to CONSORT guidelines for non-pharmacological treatments (NPT): A cross-sectional study. *BMJ Open.* 2013;3(12):1-7. doi:10.1136/bmjopen-2013-003898
116. Sarkis-Onofre R, Poletto-Neto V, Cenci MS, Pereira-Cenci T, Moher D. Impact of the CONSORT Statement endorsement in the completeness of reporting of randomized clinical trials in restorative dentistry. *J Dent.* 2017;58:54-59. doi:10.1016/j.jdent.2017.01.009
117. Caulley L, Catalá-López F, Whelan J, et al. Reporting guidelines of health research studies are frequently used inappropriately. *J Clin Epidemiol.* 2020. doi:10.1016/j.jclinepi.2020.03.006
118. Shamseer L, Hopewell S, Altman DG, Moher D, Schulz KF. Update on the endorsement of CONSORT by high impact factor journals: A survey of journal "Instructions to Authors" in 2014. *Trials.* 2016;17(1):301. doi:10.1186/s13063-016-1408-z
119. Ghosn L, Boutron I, Ravaud P. Consolidated Standards of Reporting Trials (CONSORT) extensions covered most types of randomized controlled trials, but the potential workload for authors was high. *J Clin Epidemiol.* 2019;113:168-175.
120. Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: Explanation and elaboration. *Ann Intern Med.* 2008. doi:10.7326/0003-4819-148-4-200802190-00008
121. Huwiler-Müntener K, Jüni P, Junker C, Egger M. Quality of reporting of randomized trials as a measure of methodologic quality. *J Am Med Assoc.* 2002;287(21):2801-2804. doi:10.1001/jama.287.21.2801
122. Boutron I, Altman DG, Moher D, et al. CONSORT Statement for Randomized Trials of Nonpharmacologic Treatments: A 2017 Update and a CONSORT Extension for Nonpharmacologic Trial Abstracts. *Ann Intern Med.* 2017;167(1):40-47. doi:10.7326/M17-0046
123. Hoffmann TC, Glasziou PP, Boutron I, et al. Better reporting of interventions: Template for intervention description and replication (TIDieR) checklist and guide. *BMJ.* 2014;348. doi:10.1136/bmj.g1687

124. 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;24:85-89. doi:10.1016/j.math.2016.03.004
125. Butcher NJ, Monsour A, Mew EJ, et al. Improving outcome reporting in clinical trial reports and protocols: Study protocol for the Instrument for reporting Planned Endpoints in Clinical Trials (InsPECT). *Trials.* 2019;20(1):1-10. doi:10.1186/s13063-019-3248-0
126. Butcher NJ, Mew EJ, Saeed L, et al. Guidance for reporting outcomes in clinical trials: Scoping review protocol. *BMJ Open.* 2019;9(2):1-6. doi:10.1136/bmjopen-2018-023001
127. Dwan K, Gamble C, Williamson PR, Kirkham JJ. Systematic Review of the Empirical Evidence of Study Publication Bias and Outcome Reporting Bias - An Updated Review. *PLoS One.* 2013;8(7). doi:10.1371/journal.pone.0066844
128. Hopewell S, Dutton S, Yu L-M, Chan A-W, Altman DG. The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed. *BMJ.* 2010;340. doi:10.1136/bmj.c723
129. Egger M, Smith GD. Bias in location and selection of studies. *Br Med J.* 1998;316(7124):61-66.
130. Higgins JPT, Thomas J. *Cochrane Handbook for Systematic Reviews of Interventions.* John Wiley & Sons; 2019.
131. Michaleff ZA, Costa LOP, Moseley AM, et al. CENTRAL, PEDro, PubMed, and EMBASE Are the Most Comprehensive Databases Indexing Randomized Controlled Trials of Physical Therapy Interventions. *Phys Ther.* 2011;91(2):190-197.
132. Pillastrini P, Vanti C, Curti S, et al. Using PubMed search strings for efficient retrieval of manual therapy research literature. *J Manipulative Physiol Ther.* 2015. doi:10.1016/j.jmpt.2014.11.005
133. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. *Biometrics.* 2006. doi:10.2307/2529310
134. Chen JC, Cooper RJ, McMullen ME, Schriger DL. Graph Quality in Top Medical Journals. In: *Annals of Emergency Medicine.* ; 2017. doi:10.1016/j.annemergmed.2016.08.463
135. Alvarez G, Solà I, Sitjà-Rabert M, 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. doi:10.1016/j.jclinepi.2020.01.006
136. Núñez-Cortés R, Alvarez G, Pérez-Bracchiglione J, et al. Reporting results in manual therapy clinical trials: A need for improvement. *Int J Osteopath Med.* 2021. doi:10.1016/j.ijosm.2021.06.002
137. Alvarez G, Núñez-Cortés R, Solà I, 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.* 2020;130:96-106. doi:10.1016/j.jclinepi.2020.10.018
138. Hassett AL, Williams DA. Non-pharmacological treatment of chronic widespread musculoskeletal pain. *Best Pract Res Clin Rheumatol.* 2011. doi:10.1016/j.berh.2011.01.005
139. Reimold AM, Chandran V. Nonpharmacologic therapies in spondyloarthritis. *Best Pract Res Clin Rheumatol.* 2014. doi:10.1016/j.berh.2014.10.003

140. Chang K-L, Fillingim R, Hurley RW, Schmidt S. Chronic pain management: nonpharmacological therapies for chronic pain. *FP Essent.* 2015;432:21-26. <http://www.ncbi.nlm.nih.gov/pubmed/25970869>.
141. Rubinstein SM, Van Eekelen R, Oosterhuis T, De Boer MR, Ostelo RWJGJG, Van Tulder MW. The risk of bias and sample size of trials of spinal manipulative therapy for low back and neck pain: Analysis and recommendations. *J Manipulative Physiol Ther.* 2014;37(8):523-541. doi:10.1016/j.jmpt.2014.07.007
142. Castellini G, Gianola S, Bonovas S, Moja L. Improving Power and Sample Size Calculation in Rehabilitation Trial Reports: A Methodological Assessment. *Arch Phys Med Rehabil.* 2016;97(7):1195-1201.
143. Abdul Latif L, Daud Amadera JE, Pimentel D, Pimentel T, Fregni F. Sample size calculation in physical medicine and rehabilitation: A systematic review of reporting, characteristics, and results in randomized controlled trials. *Arch Phys Med Rehabil.* 2011. doi:10.1016/j.apmr.2010.10.003
144. Lilford RJ. The ethics of underpowered clinical trials. *JAMA J Am Med Assoc.* 2002;288(17):2118-2119. <https://login.ezproxy.uniroma1.it/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=rzh&AN=119411812&lang=it&site=ehost-live&scope=site>.
145. Halpern SD, Karlawish JHT, Berlin JA. The continuing unethical conduct of underpowered clinical trials. *J Am Med Assoc.* 2002;288(3):358-362. doi:10.1001/jama.288.3.358
146. Armijo-Olivo S. The importance of determining the clinical significance of research results in physical therapy clinical research. *Brazilian J Phys Ther.* 2018;22(3):175-176. doi:10.1016/j.bjpt.2018.02.001
147. Page P. Beyond statistical significance: clinical interpretation of rehabilitation research literature. *Int J Sport Phys Ther.* 2014;9(5):726-736. <https://www.ncbi.nlm.nih.gov/pubmed/25328834>.
148. Armijo-Olivo S, Warren S, Fuentes J, Magee DJ. Clinical relevance vs. statistical significance: Using neck outcomes in patients with temporomandibular disorders as an example. *Man Ther.* 2011. doi:10.1016/j.math.2011.05.006
149. Schulz KF, Grimes DA. Sample size calculations in randomised trials: mandatory and mystical. *Lancet.* 2005;365(9467):1348-1353.
150. Fregni F, Imamura M, Chien HF, 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. doi:10.1097/PHM.0b013e3181bc0bbd
151. Hart T, Bagiella E. Design and Implementation of Clinical Trials in Rehabilitation Research. *Arch Phys Med Rehabil.* 2012;93(8, Supplement):S117--S126.
152. Hawk C, Long CR, Reiter R, Davis CS, Cambron JA, Evans R. Issues in planning a placebo-controlled trial of manual methods: results of a pilot study. *J Altern Complement Med.* 2002;8(1):21-32. doi:10.1089/107555302753507159
153. Howick J, Webster RK, Rees JL, et al. TIDieR-Placebo: A guide and checklist for reporting placebo and sham controls. *PLoS Med.* 2020;17(9):1-15. doi:10.1371/journal.pmed.1003294

154. 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(1):6-13. doi:10.2340/16501977-1071
155. Moustgaard H, Clayton GL, Jones HE, et al. Impact of blinding on estimated treatment effects in randomised clinical trials: meta-epidemiological study. *Bmj.* 2020;368:l6802-13.
156. Drucker AM, Chan A. Blindsided : challenging the dogma of masking in Fresh evidence should encourage critical thinking about when blinding really matters. 2020;229(January):1-2. doi:10.1136/bmj.m229
157. Chi-Lun-Chiao A, Chehata M, Broeker K, et al. Patients' perceptions with musculoskeletal disorders regarding their experience with healthcare providers and health services: an overview of reviews. *Arch Physiother.* 2020;10(1):1-19. doi:10.1186/s40945-020-00088-6
158. Wasmann KA, Wijsman P, Van Dieren S, Bemelman W, Buskens C. Partially randomised patient preference trials as an alternative design to randomised controlled trials: Systematic review and meta-analyses. *BMJ Open.* 2019;9(10):1-12. doi:10.1136/bmjopen-2019-031151
159. Adamson SJ, Bland JM, Hay EM, et al. Patients' preferences within randomised trials: Systematic review and patient level meta-analysis. *BMJ.* 2009;338(7686):85-87. doi:10.1136/bmj.a1864
160. Coronado RA, Bialosky JE. Manual physical therapy for chronic pain: the complex whole is greater than the sum of its parts. *J Man Manip Ther.* 2017;25(3):115-117. doi:10.1080/10669817.2017.1309344
161. Geri T, Viceconti A, Minacci M, Testa M, Rossettini G. Manual therapy: Exploiting the role of human touch. *Musculoskelet Sci Pract.* 2019;44:1-20. doi:10.1016/j.msksp.2019.07.008
162. Hróbjartsson A, Forfang E, Haahr MT, Als-Nielsen B, Brorson S. Blinded trials taken to the test: An analysis of randomized clinical trials that report tests for the success of blinding. *Int J Epidemiol.* 2007;36(3):654-663. doi:10.1093/ije/dym020
163. Sackett DL. Commentary: Measuring the success of blinding in RCTs: Don't, must, can't or needn't? *Int J Epidemiol.* 2007;36(3):664-665. doi:10.1093/ije/dym088
164. Sackett DL. Clinician-trialist rounds: 6. Testing for blindness at the end of your trial is a mug's game. *Clin Trials.* 2011;8(5):674-676. doi:10.1177/1740774511419685
165. Hróbjartsson A, Boutron I. Blinding in randomized clinical trials: Imposed impartiality. *Clin Pharmacol Ther.* 2011;90(5):732-736. doi:10.1038/clpt.2011.207
166. Hróbjartsson A. What are the main methodological problems in the estimation of placebo effects? *J Clin Epidemiol.* 2002. doi:10.1016/S0895-4356(01)00496-6
167. Foster NE, Dziedzic KS, van der Windt DAWM, Fritz JM, Hay EM. Research priorities for non-pharmacological therapies for common musculoskeletal problems: nationally and internationally agreed recommendations. *BMC Musculoskelet Disord.* 2009;10:3. doi:10.1186/1471-2474-10-3
168. Von Allmen RS, Weiss S, Tevaearai HT, et al. Completeness of follow-up determines validity of study findings: Results of a prospective repeated measures cohort study. *PLoS One.* 2015;10(10):1-13. doi:10.1371/journal.pone.0140817

169. Armijo-Olivo S, Warren S, Magee D. Intention to treat analysis, compliance, drop-outs and how to deal with missing data in clinical research: a review. *Phys Ther Rev.* 2009;14(1):36-49. doi:10.1179/174328809X405928
170. Akl EA, Briel M, You JJ, et al. Potential impact on estimated treatment effects of information lost to follow-up in randomised controlled trials (LOST-IT): systematic review. *BMJ.* 2012;344(May):1-12. doi:10.1136/bmj.e2809
171. Sedgwick P. Intention to treat analysis versus per protocol analysis of trial data. *BMJ.* 2015;350(February):1-2. doi:10.1136/bmj.h681
172. Elkins MR, Moseley AM. Intention-to-treat analysis. *J Physiother.* 2015;61(3):165-167. doi:10.1016/j.jphys.2015.05.013
173. Maitland GD, Hengeveld E, Banks K, English K. *Maitland's Vertebral Manipulation*. Elsevier Butterworth-Heinemann; 2005. <https://books.google.es/books?id=G3VsAAAA-MAAJ>.
174. Ford, Hahne, Surkitt, Chan, Richards. The Evolving Case Supporting Individualised Physiotherapy for Low Back Pain. *J Clin Med.* 2019;8(9):1334. doi:10.3390/jcm8091334
175. Ford JJ, Slater SL, Richards MC, et al. Individualised manual therapy plus guideline-based advice vs advice alone for people with clinical features of lumbar zygapophyseal joint pain: a randomised controlled trial. *Physiother (United Kingdom).* 2019;105(1):53-64. doi:10.1016/j.physio.2018.07.008
176. Ford JJ, Hahne AJ, Surkitt LD, et al. Individualised physiotherapy as an adjunct to guideline-based advice for low back disorders in primary care: A randomised controlled trial. *Br J Sports Med.* 2016;50(4):237-245. doi:10.1136/bjsports-2015-095058
177. Svedmark Å, Djupsjöbacka M, Häger C, Jull G, Björklund M. Is tailored treatment superior to non-tailored treatment for pain and disability in women with non-specific neck pain? A randomized controlled trial. *BMC Musculoskelet Disord.* 2016;17(1):1-14. doi:10.1186/s12891-016-1263-9
178. 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(1):1-13. doi:10.1186/s13104-017-2376-4
179. Lowe DB, Taylor MJ, Hill SJ. Associations between multimorbidity and additional burden for working-age adults with specific forms of musculoskeletal conditions: a cross-sectional study. *BMC Musculoskelet Disord.* 2017;18(1):1-14. doi:10.1186/s12891-017-1496-2
180. Duffield SJ, Ellis BM, Goodson N, et al. The contribution of musculoskeletal disorders in multimorbidity: Implications for practice and policy. *Best Pract Res Clin Rheumatol.* 2017;31(2):129-144.
181. Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. *J Chronic Dis.* 1967;20(8):637-648. doi:10.1016/0021-9681(67)90041-0
182. Thorpe KE, Zwarenstein M, Oxman AD, et al. A pragmatic-explanatory continuum indicator summary (PRECIS): A tool to help trial designers. *C Can Med Assoc J.* 2009;180(10). doi:10.1503/cmaj.090523
183. 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. doi:10.1136/bmj.h2147

184. Foster N, Little P. Methodological issues in pragmatic trials of complex interventions in primary care. *Br J Gen Pract.* 2012;62(594):10-11. doi:10.3399/bjgp12X616238
185. Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. *Lancet (London, England).* 2009;374(9683):86-89.
186. Glasziou P, Altman DG, Bossuyt P, et al. Reducing waste from incomplete or unusable reports of biomedical research. *Lancet (London, England).* 2014;383(9913):267-276.
187. Glasziou P, Meats E, Heneghan C, Shepperd S. What is missing from descriptions of treatment in trials and reviews? *BMJ.* 2008;336(7659):1472 LP - 1474. doi:10.1136/bmj.39590.732037.47
188. Hoffmann TC, Eructi C, Glasziou PP. Poor description of non-pharmacological interventions: Analysis of consecutive sample of randomised trials. *BMJ.* 2013;347(7924). doi:10.1136/bmj.f3755
189. Yamato TP, Maher CG, Saragiotto BT, Hoffmann TC, Moseley AM. How completely are physiotherapy interventions described in reports of randomised trials? *Physiotherapy.* 2016;102(2):121-126.
190. Juexuan C, Yuting D, Zhaoxiang B, Chi Z, Yaolong C. Better reporting of interventions: Template for intervention description and replication (TIDieR) checklist and guide. *Chinese J Evidence-Based Med.* 2020;20(12):1439-1448. doi:10.7507/1672-2531.202008182
191. Chan AW, Tetzlaff JM, Altman DG, et al. SPIRIT 2013 statement: Defining standard protocol items for clinical trials. *Ann Intern Med.* 2013. doi:10.7326/0003-4819-158-3-201302050-00583
192. Minkalis AL, Vining RD, Long CR, Hawk C, de Luca K. A systematic review of thrust manipulation combined with one conservative intervention for rotator cuff and related non-surgical shoulder conditions. *J Can Chiropr Assoc.* 2018;62(1):5-17. <https://www.ncbi.nlm.nih.gov/pubmed/30270924>.
193. Zavala-González J, Pavez-Baeza F, Gutiérrez-Espinoza H, Olguín-Huerta C. The effectiveness of joint mobilization techniques for range of motion in adult patients with primary adhesive capsulitis of the shoulder: a systematic review and meta-analysis. *Medwave.* 2018;18(5):e7265. doi:10.5867/medwave.2018.05.7265
194. Carnes D, Plunkett A, Ellwood J, Miles C. Manual therapy for unsettled, distressed and excessively crying infants: a systematic review and meta-analyses. *BMJ Open.* 2018;8(1):e019040. doi:10.1136/bmjopen-2017-019040
195. de Luca KE, Fang SH, Ong J, Shin K-S, Woods S, Tuchin PJ. The Effectiveness and Safety of Manual Therapy on Pain and Disability in Older Persons With Chronic Low Back Pain: A Systematic Review. *J Manip Physiol Ther.* 2017;40(7):527-534. doi:10.1016/j.jmpt.2017.06.008
196. Bizzarri P, Buzzatti L, Cattrysse E, Scafoglieri A. Thoracic manual therapy is not more effective than placebo thoracic manual therapy in patients with shoulder dysfunctions: A systematic review with meta-analysis. *Musculoskelet Sci Pract.* 2018;33:1-10. doi:10.1016/j.msksp.2017.10.006
197. Laimi K, Mäkilä A, Bärlund E, et al. Effectiveness of myofascial release in treatment of chronic musculoskeletal pain: a systematic review. *Clin Rehabil.* 2018;32(4):440-450. doi:10.1177/0269215517732820

198. Wu J, Wang Y, Wang Z. The effectiveness of massage and touch on behavioural and psychological symptoms of dementia: A quantitative systematic review and meta-analysis. *J Adv Nurs*. 2017;73(10):2283-2295. doi:10.1111/jan.13311
199. Smith CA, Levett KM, Collins CT, Dahlen HG, Ee CC, Suganuma M. Massage, reflexology and other manual methods for pain management in labour. *Cochrane Database Syst Rev*. 2018;3:CD009290. doi:10.1002/14651858.CD009290.pub3
200. Furlan AD, Giraldo M, Baskwill A, Irvin E, Imamura M. Massage for low-back pain. *Cochrane Database Syst Rev*. 2015. doi:10.1002/14651858.cd001929.pub3
201. Westad K, Tjoestolsen F, Hebron C. The effectiveness of Mulligan's mobilisation with movement (MWM) on peripheral joints in musculoskeletal (MSK) conditions: A systematic review. *Musculoskelet Sci Pract*. 2018. doi:10.1016/j.msksp.2018.12.001
202. Lee N-W, Kim G-H, Heo I, et al. Chuna (or Tuina) Manual Therapy for Musculoskeletal Disorders: A Systematic Review and Meta-Analysis of Randomized Controlled Trials. *Evid Based Complement Altern Med*. 2017;2017:8218139. doi:10.1155/2017/8218139
203. Xu Q, Chen B, Wang Y, et al. The Effectiveness of Manual Therapy for Relieving Pain, Stiffness, and Dysfunction in Knee Osteoarthritis: A Systematic Review and Meta-Analysis. *Pain Physician*. 2017;20(4):229-243. <https://www.ncbi.nlm.nih.gov/pubmed/28535547>.
204. Mischke JJ, Jayaseelan DJ, Sault JD, Emerson Kavchak AJ. The symptomatic and functional effects of manual physical therapy on plantar heel pain: a systematic review. *J Man Manip Ther*. 2017;25(1):3-10. doi:10.1080/10669817.2015.1106818
205. Krul M, van der Wouden JC, Kruithof EJ, van Suijlekom-Smit LWA, Koes BW. Manipulative interventions for reducing pulled elbow in young children. *Cochrane Database Syst Rev*. 2017;(7). doi:10.1002/14651858.CD007759.pub4
206. Gross A, Langevin P, Burnie SJ, et al. Manipulation and mobilisation for neck pain contrasted against an inactive control or another active treatment. *Cochrane Database Syst Rev*. 2015. doi:10.1002/14651858.cd004249.pub4
207. Silva TM, Costa LCM, Costa LOP. Evidence-Based Practice: a survey regarding behavior, knowledge, skills, resources, opinions and perceived barriers of Brazilian physical therapists from São Paulo state. *Braz J Phys Ther*. 2015;19(4):294-303. doi:10.1590/bjpt-rbf.2014.0102
208. da Silva TM, Costa L da CM, Garcia AN, Costa LOP. What do physical therapists think about evidence-based practice? A systematic review. *Man Ther*. 2015;20(3):388-401. doi:10.1016/j.math.2014.10.009
209. Altman DG, Schulz KF, Moher D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med*. 2001;134(8):663-694. <https://www.ncbi.nlm.nih.gov/pubmed/11304107>.
210. Vetter TR, Mascha EJ. Defining the Primary Outcomes and Justifying Secondary Outcomes of a Study: Usually, the Fewer, the Better. *Anesth Analg*. 2017;125(2):678-681. doi:10.1213/ANE.0000000000002224
211. Bialosky J, Bishop M, Price D, Robinson M, George S. Mechanisms of manual therapy. *Mech Man Ther*. 2008;14(5):103 p-103 p. doi:10.1016/j.math.2008.09.001.The
212. Kamper SJ. Interpreting outcomes 1-Change and difference: Linking evidence to

- practice. *J Orthop Sports Phys Ther.* 2019;49(5):357-358. doi:10.2519/jospt.2019.0703
213. Goodman S. A Dirty Dozen: Twelve P-Value Misconceptions. *Semin Hematol.* 2008;45(3):135-140. doi:10.1053/j.seminhematol.2008.04.003
214. Greenland S, Senn SJ, Rothman KJ, et al. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *Eur J Epidemiol.* 2016;31(4):337-350. doi:10.1007/s10654-016-0149-3
215. Bassler D, Briel M, Montori VM, et al. Stopping randomized trials early for benefit and estimation of treatment effects: systematic review and meta-regression analysis. *JAMA.* 2010;303(12):1180-1187. doi:10.1001/jama.2010.310
216. Bleakley C, Reijgers J, Smoliga JM. Many High-Quality Randomized Controlled Trials in Sports Physical Therapy Are Making False-Positive Claims of Treatment Effect: A Systematic Survey. *J Orthop Sport Phys Ther.* 2020;50(2):104-109. doi:10.2519/jospt.2020.9264
217. Kamper SJ. Confidence intervals: Linking evidence to practice. *J Orthop Sports Phys Ther.* 2019;49(10):763-764. doi:10.2519/jospt.2019.0706
218. Hefford C, Abbott JH, Baxter GD, Arnold R. Outcome measurement in clinical practice: practical and theoretical issues for health related quality of life (HRQOL) questionnaires. *Phys Ther Rev.* 2013;16(3):155-167.
219. Gagnier JJ, Moher D, Boon H, Beyene J, Bombardier C. Investigating clinical heterogeneity in systematic reviews: a methodologic review of guidance in the literature. *BMC Med Res Methodol.* 2012;12(1):1.
220. Smith T. The challenges of heterogeneity across physiotherapy trials when considering meta-analysis. *Phys Ther Rev.* 2013;19(1):39-40.
221. 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(2):119-125.
222. Calvert M, Blazeby J, Altman DG, et al. Reporting of Patient-Reported Outcomes in Randomized Trials: The CONSORT PRO Extension. *JAMA* 2013;309(8):814-822.
223. Pocock S, Trivison T, Wruck L. How to interpret figures in reports of clinical trials. *Bmj.* 2008;336(7654):1166-1169. doi:10.1136/bmj.39561.548924.94
224. Srivastava R, Cleveland WS. *The Elements of Graphing Data.* Vol 24.; 1987. doi:10.2307/3151402
225. Puhan MA, ter Riet G, Eichler K, Steurer J, Bachmann LM. More medical journals should inform their contributors about three key principles of graph construction. *J Clin Epidemiol.* 2006;59(10):1017-1022. doi:10.1016/j.jclinepi.2005.12.016
226. Pocock SJ, Trivison TG, Wruck LM. Figures in clinical trial reports: current practice & scope for improvement. *Trials.* 2007;8:36. doi:10.1186/1745-6215-8-36
227. Schriger DL, Cooper RJ. Achieving graphical excellence: suggestions and methods for creating high-quality visual displays of experimental data. *Ann Emerg Med.* 2001;37(1):75-87. doi:10.1067/mem.2001.111570
228. Schriger DL, Sinha R, Schroter S, Liu PY, Altman DG. From submission to publication: a retrospective review of the tables and figures in a cohort of randomized controlled trials submitted to the British Medical Journal. *Ann Emerg Med.* 2006;48(6):750-756,

- 756.e1-21. doi:10.1016/j.annemergmed.2006.06.017
229. Karpouzis F, Bonello R, Pribicevic M, Kalamir A, Brown BT. Quality of reporting of randomised controlled trials in chiropractic using the CONSORT checklist. *Chiropr Man Therap.* 2016;24(1):19. doi:10.1186/s12998-016-0099-6
 230. Turner L, Shamseer L, Altman DG, Schulz KF, Moher D. Does use of the CONSORT Statement impact the completeness of reporting of randomised controlled trials published in medical journals? A Cochrane review. *Syst Rev.* 2012;1(1):1. doi:10.1186/2046-4053-1-60
 231. Blanco D, Altman D, Moher D, Boutron I, Kirkham JJ, Cobo E. Scoping review on interventions to improve adherence to reporting guidelines in health research. *BMJ Open.* 2019;9(5):e026589--12.
 232. Hopewell S, Boutron I, Altman DG, et al. Impact of a web-based tool (WebCONSORT) to improve the reporting of randomised trials: results of a randomised controlled trial. *BMC Med.* 2016;14(1):199.
 233. Pandis N, Turpin DL. Enhancing CONSORT compliance for improved reporting of randomized controlled trials. *Am J Orthod Dentofac Orthop.* 2014. doi:10.1016/j.ajodo.2013.11.005
 234. Barnes C, Boutron I, Giraudeau B, Porcher R, Altman DG, Ravaud P. Impact of an online writing aid tool for writing a randomized trial report: The COBWEB (Consort-based WEB tool) randomized controlled trial. *BMC Med.* 2015. doi:10.1186/s12916-015-0460-y
 235. Blanco D, Biggane AM, Cobo E. Are CONSORT checklists submitted by authors adequately reflecting what information is actually reported in published papers? January 2018:1-4.
 236. Moher D, Altman DG. Four Proposals to Help Improve the Medical Research Literature. *PLoS Med.* 2015. doi:10.1371/journal.pmed.1001864
 237. Moher D, Jones A, Lepage L. Use of the CONSORT statement and quality of reports of randomized trials. *J Am Med Assoc.* 2001;285(15):1992-1995. doi:10.1001/jama.285.15.1992
 238. Negrini S, Arienti C, Pollet J, et al. Clinical replicability of rehabilitation interventions in randomized controlled trials reported in main journals is inadequate. *J Clin Epidemiol.* 2019;114:108-117.
 239. Savithra P, Nagesh LS. Have CONSORT guidelines improved the quality of reporting of randomised controlled trials published in public health dentistry journals? *Oral Health Prev Dent.* 2013;11(2):95-103. doi:10.3290/j.ohpd.a29359
 240. Macpherson H, Altman DG, Hammerschlag R, Youping L, Taixiang W. Revised STAndards for Reporting Interventions in Clinical Trials of Acupuncture (STRICTA): Extending the CONSORT Statement. 2010;7(6). doi:10.1371/journal.pmed.1000261
 241. Asher GN, Motsinger-Reif AA, Jonas DE, Viera AJ. Quality of reporting on randomised controlled trials of auriculotherapy for pain. *Acupunct Med.* 2011;29(2):122-126.
 242. Svenkerud S, MacPherson H. The impact of STRICTA and CONSORT on reporting of randomised control trials of acupuncture: a systematic methodological evaluation.

- Acupunct Med.* September 2018. doi:10.1136/acupmed-2017-011519
243. Kim KH, Kang JW, Lee MS, Lee J-D. Assessment of the quality of reporting in randomised controlled trials of acupuncture in the Korean literature using the CONSORT statement and STRICTA guidelines. *BMJ Open.* 2014;4(7):e005068--e005068.
244. Jia P, Tang L, Yu J, Liu J, Kang D, Sun X. The quality of reporting in randomized controlled trials of acupuncture for knee osteoarthritis: A cross-sectional survey. *PLoS One.* 2018;13(4):e0195652--14.
245. Groeneweg R, Rubinstein SM, Oostendorp RAB, Ostelo RWJG, Van Tulder MW. Guideline for Reporting Interventions on Spinal Manipulative Therapy: Consensus on Interventions Reporting Criteria List for Spinal Manipulative Therapy (CIRCLE SMT). *J Manip Physiol Ther.* 2017;40(2):61-70.
246. Page P, Hoogenboom B, Voight M. Improving the Reporting of Therapeutic Exercise Interventions in Rehabilitation Research. *Int J Sports Phys Ther.* 2017;12(2):297-304. <http://www.ncbi.nlm.nih.gov/pubmed/28515984><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5380872>.
247. Major DH, Røe Y, Grotle M, et al. Content reporting of exercise interventions in rotator cuff disease trials: results from application of the Consensus on Exercise Reporting Template (CERT). *BMJ open Sport Exerc Med.* 2019;5(1):e000656. doi:10.1136/bmjsem-2019-000656
248. Slade SC, Finnegan S, Dionne CE, Underwood M, Buchbinder R. The Consensus on Exercise Reporting Template (CERT) applied to exercise interventions in musculoskeletal trials demonstrated good rater agreement and incomplete reporting. *J Clin Epidemiol.* 2018;103:120-130. doi:10.1016/j.jclinepi.2018.07.009
249. Keene DJ, Forde C, Sugavanam T, Williams MA, Lamb SE. Exercise for people with a fragility fracture of the pelvis or lower limb: a systematic review of interventions evaluated in clinical trials and reporting quality. *BMC Musculoskelet Disord.* 2020;21(1):435. doi:10.1186/s12891-020-03361-8
250. Kattackal T-R, Cavallo S, Brosseau L, et al. Assessing the reporting quality of physical activity programs in randomized controlled trials for the management of juvenile idiopathic arthritis using three standardized assessment tools. *Pediatr Rheumatol Online J.* 2020;18(1):41. doi:10.1186/s12969-020-00434-9
251. Kent P, O'Sullivan PB, Keating J, Slade SC. Evidence-based exercise prescription is facilitated by the Consensus on Exercise Reporting Template (CERT). *Br J Sports Med.* 2018;52(3):147 LP - 148. doi:10.1136/bjsports-2016-097405
252. Uttley L, Montgomery P. The influence of the team in conducting a systematic review. *Syst Rev.* 2017;6(1):4-7. doi:10.1186/s13643-017-0548-x
253. Eden, Laura Levit, Alfred Berg and SM. *Finding What Works in Health Care: Standards for Systematic Reviews.* (Eden J, Levit L, Berg A, Morton S, eds.). Washington, DC: The National Academies Press; 2011. doi:10.17226/13059
254. Rücker G. A two-stage trial design for testing treatment, self-selection and treatment preference effects. *Stat Med.* 1989;8(4):477-485. doi:10.1002/sim.4780080411

TESIS DOCTORAL

Retos metodológicos en la evaluación de la eficacia de la terapia manual

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

Gerard Alvarez Bustins

Directores: Dr. Gerard Urrútia Cuchí y Dr. Xavier Bonfill Cosp

Octubre 2021



Universitat Autònoma de Barcelona