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

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

[a] PhD student in Biomedical Research Methodology and Public Health in the Medical Department of the Universitat Autonoma de Barcelona. Barcelona, Spain
[b] Iberoamerican Cochrane Centre – Biomedical Research Institute Sant Pau, IIB Sant Pau, Barcelona, Spain
[c]Department of Physical Therapy, Faculty of Health Science Blanquerna, Ramon Llull University, Barcelona, Spain.
[d] Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, London, United Kingdom
[e] UCO School of Osteopathy, Health Sciences University, London, United Kingdom
[f] Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP), Madrid, Spain
[g] Universitat Autònoma de Barcelona, Bellaterra, Spain
[h] Department of Physical Therapy, Faculty of Medicine, University of Chile, Santiago, Chile.
[i] Interdisciplinary Centre for Health Studies (CIESAL), Universidad de Valparaíso, Viña del Mar, Chile
[j] Global Research on Wellbeing (GRoW) Research Group, Ramon Llull University, Barcelona, Spain
[k]Tecnocampus, Universitat Pompeu Fabra, Mataró-Maresme, Barcelona, Spain
[l] Institut ANALGESIA, Faculté de Médecine, Université Clermont Auvergne, Clermont-Ferrand, France
[m] CHU Clermont-Ferrand, Inserm 1107 Neuro-Dol, Service de pharmacologie médicale, Université Clermont Auvergne, Clermont-Ferrand, France
[n] Institut d'Investigació Biomèdica Sant Pau (IIB SANT PAU), Barcelona, Spain

*Corresponding author: Iberoamerican Cochrane Center, C. Sant Antoni Maria Claret 167, Pavelló 18, Planta 008025 Barcelona, Catalunya, Spain Tel.: +34 620 214 206 E-mail address: gerardalv@gmail.com

KEYWORDS

Effectiveness Clinical Trials, Musculoskeletal Manipulations, Randomised Controlled Trials, manual therapy, generalizability, PRECIS-2

ABSTRACT

Objectives: The study aimed to assess the design, reporting, and risk of bias in effectiveness trials in manual therapy, comparing pragmatic with non-pragmatic trials and trials with and without placebo controls.

Study Design and Setting: We searched MEDLINE and the Cochrane Central Register of Controlled Trials for randomised controlled trials with the term 'effectiveness' in the title or abstract in the field of manual therapy from inception to January 2024. Two independent reviewers extracted data on specific study characteristics, their reporting

and risk of bias and assessed them using the PRECIS-2 tool. Descriptive analysis using frequencies and percentages and a relation analysis between PRECIS-2 scores and specific study characteristics were performed.

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

Conclusion: The term effectiveness is used in MT trials independently of its pragmatic connotations. Using a placebo or active control does not modify the pragmatic attitude of the effectiveness trials.

HIGHLIGHTS

Key findings

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

What this adds to what is known related to methods research within the field of clinical epidemiology

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

What is the implication, what should change now

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

1. INTRODUCTION

Both efficacy and effectiveness randomised clinical trials (RCTs) are essential for evaluating interventions [1,2]. Efficacy trials are conducted under controlled conditions to explore the causal relationship between interventions and their physiological effects (explanatory attitude) [3]. Conversely, effectiveness trials aim to inform healthcare professionals by comparing interventions under real-world conditions (pragmatic attitude) [3,4]. The terms commonly used in the literature to designate an article as explanatory are efficacy trials and “pragmatic”, or “naturalistic”, to designate them as pragmatic, or referring to the “effectiveness” of an intervention [3,5–9]. These trial types are better understood as a continuum rather than a strict dichotomy, as they often share overlapping methodologies and conceptual frameworks [4,10–13]. Further, the terms "attitude" and "study intent” must be differentiated. The chosen “attitude” is the overarching orientation of the research, either choosing a more pragmatic or a more efficacy-focused orientation. In contrast, the “study intent” is the goal pursued within that orientation, that is, how the author states and justifies the specific objective or research question [3,10], such as ”determining if a specific MT technique is correlated with a specific heart rate variability change (explanatory attitude)” versus, “evaluating if a adding MT to the standard treatment for obstructive pulmonary infection outperforms the

standard treatment alone in real-world conditions (pragmatic attitude)". Effectiveness trials are necessary because they focus on increasing research's relevance and applicability to clinical practice and policy decision-making. Thus, they might ensure external validity but also employ methods that enhance internal trial validity, such as randomisation and treatment fidelity monitoring.

In manual therapy (MT), efficacy and effectiveness trials are well-documented [1,14]. However, concerns have been raised regarding their overall methodological rigor due to internal validity issues and reporting deficiencies [15–23]. Although the term "effectiveness" is commonly used in MT trials [24–27], a systematic review revealed a lack of pragmatism in many cases [1]. This has led to concerns about whether these trials accurately simulate clinical practice and deliver meaningful insights for healthcare decisions [1,28,29]. The inconsistent use of terminology has implications for research, potentially affecting the interpretation of trial outcomes and influencing clinical guideline recommendations [30,31].

Research shows that MT trials often lean towards explanatory designs rather than pragmatic approaches [1]. Previous findings from the author group confirmed that self-labelled MT pragmatic RCTs (note that to facilitate readability, those will be named pragmatic trials -PT- throughout the manuscript) have not increased in recent years, unlike in other disciplines [1,10,32–34]. Furthermore, their pragmatic attitude remains moderate, and better internal validity and reporting are needed [34]. The misuse of "efficacy" and "effectiveness" as interchangeable terms has resulted in research designs often inconsistent with their stated objectives [4,31,35,36]. This issue is particularly prevalent in studies focused on the clinical effects of interventions with limited efficacy evidence but widespread clinical use, such as MT [32,37].

Given the conceptual uncertainty regarding terminology, this study aimed to compare effectiveness MT trials self-labelled as pragmatic with those that are not. A substantial body of research exists that neither applies well in real-world settings nor is designed with the internal validity required to be considered efficacy studies. By assessing the methodological differences, the researchers sought to clarify whether there are critical design disparities and promote more accurate reporting and understanding of trial characteristics. This research underscores the need for trial design and terminology clarity to strengthen academic inquiry and clinical guideline development. Establishing consistent definitions and methodological practices will help improve the quality and reliability of MT trials, ultimately benefiting clinical practice and healthcare decision-making.

2. METHODS

2.1 Protocol registration

We conducted a methodological review of effectiveness trials on MT interventions, following the guidelines of the Cochrane Handbook for systematic reviews of interventions [38] and reported according to the PRISMA guidelines [39] (Supplementary file 1). The study was prospectively registered on the Open Science Framework (DOI 10.17605/OSF.IO/WKEPZ).

2.2 Eligibility criteria

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

2.3 Search strategy

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

2.4 Study selection

The records were uploaded onto the Rayyan software (www.rayyan.ai) [40]. Upon deduplication, the references were screened by two independent reviewers (SR, GA) based on the title and summary, resolving disagreements through discussion.

The references were initially classified into two groups: GROUP 1 consisting of effectiveness trials labelled as "pragmatic" in the title and/or abstract (PT), and GROUP

2 comprising effectiveness trials that were not self-identified as pragmatic (ET). For further analysis, we split the ET group into studies that utilised a placebo control group (ETpl) - GROUP 2A- and those that employed an accepted or established comparator (ETac)- GROUP 2B, based on title/abstract information. Resulting in three groups (GROUPS 1, 2A and 2B). we determined the sample of studies to be included in each group according to the PT found. As we found many more ET than PT, we randomised the effectiveness trials by matching their number to the number of PTs found.

2.5 Data collection process

Two reviewers, SR and an additional member from GA, RN, JB, DH, CF, and JP, gathered data from the included studies and resolved disagreements through consensus. The involvement of a third party was not required to resolve any discrepancies.

Before the review process, the team met with Dr Kirsty Loudon, the developer of the PRECIS-2 tool [41]. It was deemed suitable to involve a member of the PRECIS-2 tool authors, given that PRECIS-2 was conceptualised as a tool for trial designers rather than as a method for retrospective assessment of pragmatism. The objective was to ensure that the method was utilised in the most reliable manner. Three articles were examined, compared, and discussed against Dr Loudon's criteria.

2.7. Data items

Three authors (SR, GA, GU) designed a data extraction form. It included bibliometric identification elements, the intent of the trial, the rationale of the intervention given by the authors, the experimental and control interventions, the limitations reported by the authors, the PRECIS-2 tool assessment, the CONSORT reporting assessment and the RISK of BIAS (RoB) assessment. Information about the data extraction form and the methods used to assess each item are included in Supplementary File 3 and reported elsewhere [34].

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

2.8 Data analysis

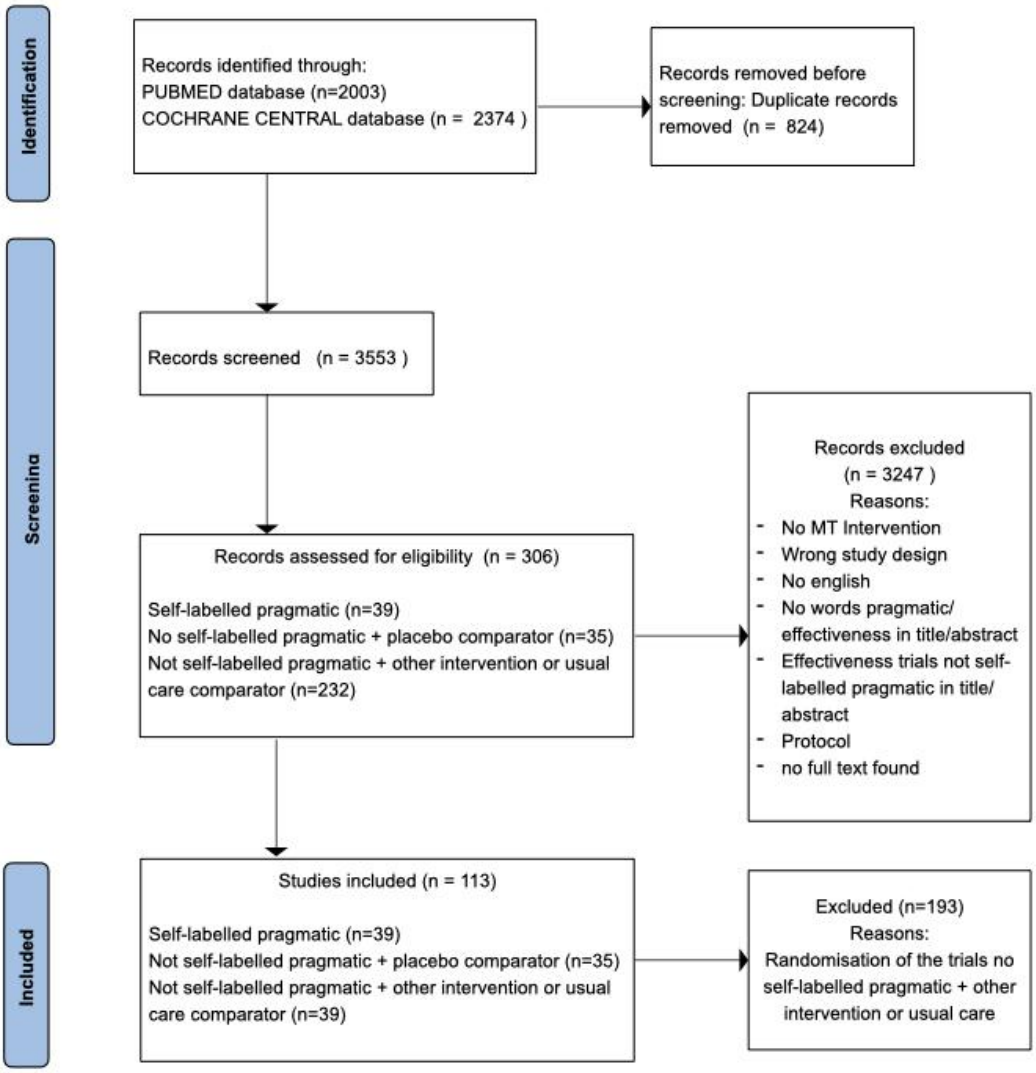
A descriptive analysis of the categorical variables was conducted, wherein results were presented as relative and absolute frequencies. Likewise, means and standard deviations were used to describe quantitative variables. Median and interquartile ranges were applied for ordinal variables. Meanwhile, the PRECIS-2 wheel was used to present the pragmatic attitude of each article graphically. Further, variables were compared using ANOVA for quantitative variables, Chi-square for categorical variables and Krustall Wallis for ordinal variables. According to previous research [42–44], we employed a 9-45 point scale to evaluate the total score PRECIS-2, apart from the PRECIS average. Ratings of 9-22 were described as slightly pragmatic, 23-34 moderately pragmatic, and $\geqslant$35 very pragmatic. We conducted post-hoc analysis utilising the Scheffé test for ANOVA and the Mann-Whitney test for Krustall Wallis. The Spearman test was performed to analyse the correlation between PRECIS-2 and some descriptive characteristics. The significance level was set at 5% (alpha=0.05). All data were analysed utilising IBM-SPSS software (V26.0).

3. RESULTS

3.1 Study selection

The search identified 3553 articles. After screening the title, abstract, and full-text disponibility, the final sample comprised 306 studies. Among these, 39 were identified as self-labelled pragmatic in the title and/or abstract, whereas 267 were 'effectiveness' Randomized Controlled Trials not self-labelled as pragmatic (ET). We did not identify any articles in TM with the term naturalistic. Of the 267 effectiveness trials, 35 were placebo-controlled trials (ETpl), while the remaining 232 employed other accepted or established interventions (ETac). To balance and compare the three groups, the 232 ETac were randomised using the Microsoft Excel random generator to get a group of 39 trials (the same number of trials as in the PT group). Resulting in a total sample for this review of 113 studies. The PRISMA diagram is shown in Figure 1.

Figure 1. PRISMA flow diagram



246
247 3.2 Study characteristics of the sample
248
249 Table 1 summarises the main characteristics of our sample. A detailed table with all
250 included studies and some additional characteristics can be found in Supplementary File
251 4. The average number of participants was significantly higher in pragmatic trials (mean

= 169) compared to ETac (mean = 65) and ETpl (mean = 70). In PT, the most frequent intervention studied was a non-protocolised combination of techniques (54%;21/39), whereas in ETac, it was a protocol of a combination of techniques (51%;20/39) and a protocol of an isolated technique in 49% (17/35) of the ETpl. No studies blinded therapists. Blinding of external assessors was more common in ETpl (74%) compared to PT (69%) and ETac (59%). Multicentric settings were more prevalent in pragmatic trials (56%), whereas effectiveness trials primarily used unicentric settings. The publication tendency of each type of design until 2024 is shown in Figure 2.

Table 1. Main characteristics of the sample

		PT (N=39)	ETac (N=39)	ETpl (N=35)
Number of participants (Mean)		169 (SD=151))	65 (SD=58)	70 (SD= 44)
N of participants			Per cent (n**)	
INTERVENTION				
Combination of non-protocolised techniques		54% (21)	10% (4)	9% (3)
Protocol of a combination of techniques		13% (5)	51% (20)	23% (8)
Isolated non-protocolised technique		15% (6)	8% (3)	8% (3)
Protocol of an isolated technique		5% (2)	18% (7)	49% (17)
Combination of non-protocolised therapies		10% (4)	3% (1)	3% (1)
Protocol of a combination of therapies		3% (1)	10% (4)	9% (3)
CONTROL INTERVENTION				
2 arms	test treatment vs other active intervention	31% (12)	46% (18)	0% (0)
	test treatment vs placebo	3% (1)	0% (0)	69% (24)
	test treatment vs usual care	33% (13)	38% (15)	0% (0)
	test treatment vs no intervention	13% (5)	5% (2)	0% (0)
3 arms	test treatment vs 2 other active interventions	5% (2)	3% (1)	0% (0)
	test treatment vs 1 other active intervention and 1 placebo	2% (1)	3% (1)	11% (4)
	test treatment vs 1 active intervention and 1 usual care	10% (4)	3% (1)	0% (0)

	test treatment vs 1 usual care and 1 placebo	2% (1)	0% (0)	17% (6)
	test treatment vs 1 other active intervention and 1 no intervention	2% (1)	0% (0)	0% (0)
4 arms	test treatment vs 1 other active intervention, 1 no intervention and 1 placebo	0% (0)	3% (1)	3% (1)
BLINDING (YES)				
Participants		8% (3)	8% (3)	20% (7)
Therapists		0% (0)	0% (0)	0% (0)
External assessors to providers and patients		69% (27)	59% (23)	74% (26)
Statistician		43% (17)	18% (7)	20% (7)
RATIONALE				
yes (rationale provided)		**64% (25)**	**64% (25)**	**63% (22)**
Comparative effectiveness data		61%	38%	31%
Mechanistic experiments		36%	54%	63%
Highly sham-controlled		3%	8%	6%
no (no rationale provided)		**21% (8)**	**28% (11)**	**26% (9)**
unclear		**13% (5)**	**8% (3)**	**11% (4)**
FUNDING				
yes		67% (26)	21% (8)	37% (13)
FOLLOW-UP				
No follow-up		18% (7)	23% (9)	28% (9)
< 2 weeks		0% (0)	13% (5)	3% (1)
2-4 weeks		2% (1)	8% (3)	3% (1)
4-12 weeks		10% (4)	20% (8)	34% (12)
3-6 months		18% (7)	15% (6)	20% (7)
6-12 months		20% (8)	15% (6)	9% (3)

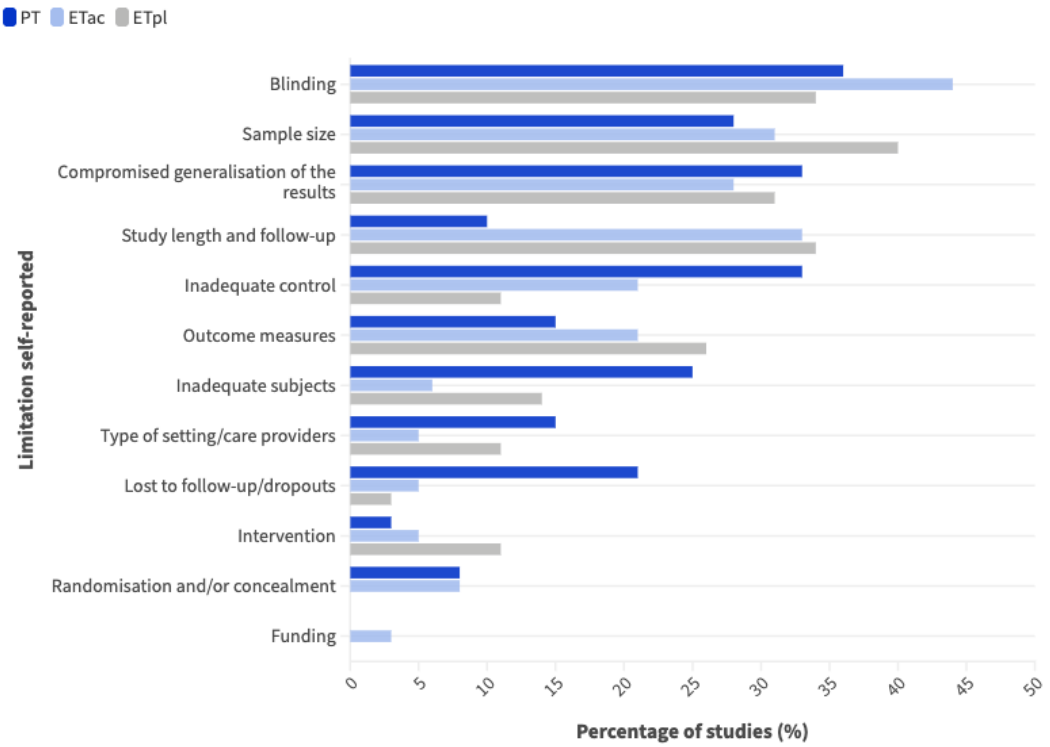
>1 year	26% (10)	5% (2)	3% (1)
*individualised	5% (2)	0 % (0)	0 % (0)
SETTING			
Multicentric	56% (22)	8% (3)	20% (7)
Unicentric	33% (13)	87% (34)	77% (27)
Unclear reported	10% (4)	5% (2)	3% (1)

*trials assessing how many weeks patients remained pain-free

** N of studies =113

Specific characteristics of pragmatic trials were reported elsewhere [34,45]. Although "Pragmatic" was not mentioned in titles or abstracts, 64% of ETac studies claimed a pragmatic approach, compared to 31% of ETpl. Common self-reported study limitations included participant blinding (14/17/12), generalisability (13/11/11), and sample size (11/12/14). Inadequate control was frequent in PT (13) and ETac (8) but rare in ETpl (4). Study length limitations were prevalent in effectiveness trials. Figure 2 illustrates these limitations across study types.

Figure 2. Comparison of limitations reported between studies



3.3 PRECIS Assessment

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

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

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

Figure 3. PRECIS-2 wheels of each study's type

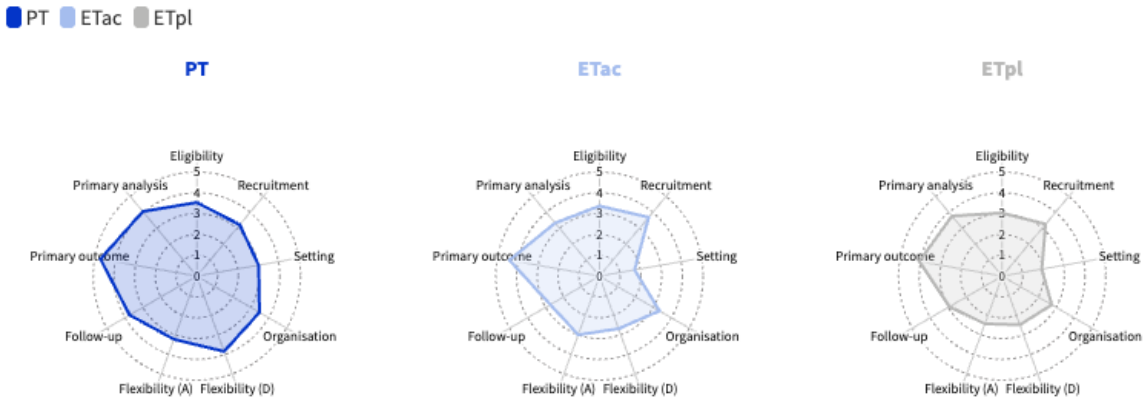


Table 3 shows PRECIS-2 rates by study type. Most domains in pragmatic trials, except Domain 6 (adherence), were pragmatic in over 50% of studies. For effectiveness trials, Domains 1 (eligibility), 2 (recruitment), and 8 (outcome measure) were pragmatic in more than 50% of cases. Domain 3 (setting) was largely explanatory in ET (82%), while Domain 5 (flexibility) was pragmatic in over 50% of PT but explanatory in most ET.

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

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

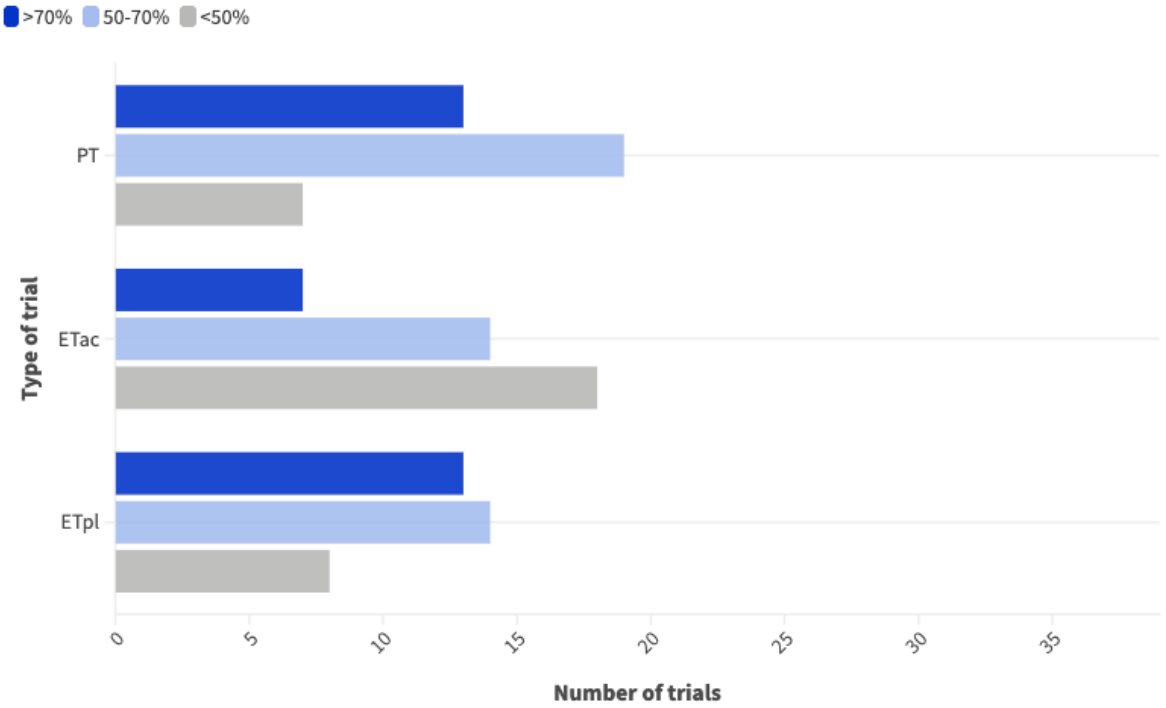
PT: pragmatic trials; ETac: Effectiveness trials active comparator; ETpl: Effectiveness trials placebo comparator; Blank: percentage of studies without information about the domain within the study

3.4 CONSORT Assessment

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

One in four trials reported details related to more than 70% of CONSORT items (13/39 of the PT, 9/39 of the ETac and 13/32 of the ETpl), between 50% to 70% in 41% of the trials (46/113) and lower than 50% of the CONSORT items in 28% of the sample (32/113). Figure 4 shows the percentage of compliance with CONSORT items depending on the study type.

Figure 4. Compliance with CONSORT items



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

Intervention descriptions were adequate in over 80% of studies (95/113). Still, the authors did not sufficiently report details of the intervention, the standardisation or individualisation of the intervention and the additional sources to resemble clinical

practice. Item 5b (Description of the different components of the interventions and, when applicable, description of the procedure for tailoring the interventions to individual participants) was better reported in the ETpl group (77%) than the ETac (56%) and the PT groups (54%). Moreover, less than 11% (12/113) of the trials reported whether and how the adherence of participants or care providers was assessed.

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

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

3. 5 Risk Of Bias

Selection bias (random sequence generation and allocation concealment) was lower in pragmatic trials than in effectiveness trials. Detection bias was higher in PT and ETac than ETpl. A table comparing the RoB between types of studies is presented in Supplementary File 7.

Please refer to Supplementary File 8 to compare the studies with a low vs. high/unclear RoB according to compliance with CONSORT items. The results showed that the percentage of compliance with CONSORT items was higher in the trials with low RoB in the following domains: random sequence generation, allocation concealment, incomplete outcome data, and selective reporting.

4. DISCUSSION

This review compared the design, reporting, and bias risk of manual therapy (MT) effectiveness RCTs, comparing trials self-labeled as pragmatic or not by the study authors, and using a placebo control or an active comparator. The results showed a moderate pragmatic attitude regardless of labelling or placebo use. Although many trials aim to emulate clinical practice, designs often lack pragmatic methodology. The review also found that reporting was generally inadequate, especially when not self-labelled pragmatic or when using usual care/other interventions instead of placebo controls as comparators. The findings highlight areas for improvement in designing and reporting MT

effectiveness RCTs to better align with clinical practice and enhance methodological rigour.

Designing MT effectiveness RCTs and correct use of the terminology

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

Pragmatic attitude and methodological characteristics of effectiveness MT RCTs

Effectiveness and pragmatic trials are terms that can be used interchangeably in many contexts to refer to clinical trials designed to answer whether an intervention works under routine health care conditions [9]. However, effectiveness trials scored lower on PRECIS-2 (2.8) than self-labelled pragmatic trials (3.5), suggesting a greater emphasis on pragmatic design when explicitly labelled. In our sample, in contrast to PT, effectiveness trials were primarily explanatory in the setting and intervention domains and had smaller sample sizes.

Experimental Intervention. The most common interventions in effectiveness trials were protocolised combinations of techniques (ETac) and isolated techniques (ETpl), placing these studies closer to the efficacy pole of the continuum. Adopting non-protocolised, patient-centred interventions could better reflect clinical practice [37,46,47]. In non-pharmacological research, such approaches are essential for generating relevant and actionable clinical results [1]. However, the lack of standardised interventions poses a significant challenge. Guidelines such as TIDIER [48] are essential for effective reporting and replicability, including detailed annexes to avoid interpretative gaps.

Additionally, the use of placebo comparators, unicentric settings, and blinding, which is common in explanatory studies, has led to a debate on their pragmatism [46–50].

Whether the same discussion should pertain to non-labelled pragmatic effectiveness trials is debatable, as "effectiveness" is often deemed synonymous with "pragmatic" in research contexts [5–8].

Placebo control. Including placebo control interventions in MT RCTs may clarify therapeutic effects, which are challenging to isolate in efficacy studies. The ongoing PRECIS-3 update aims to assess control groups more effectively, highlighting the need for further research on the role of placebos in PT, particularly in MT studies where efficacy trials struggle to pinpoint mechanisms despite demonstrating clinical outcomes.

Non-pharmacological interventions differ from pharmacological ones in development and implementation, often lacking prior efficacy studies. Our review found no significant variations in the PRECIS-2 scores between trials using placebo or other comparators, indicating a moderately pragmatic approach. This finding is consistent with the analysis conducted by Devos et al. on nursing interventions [42]. Dal-Re et al. argue that placebo-controlled trials cannot replicate clinical practice and should not be considered pragmatic. However, the PRECIS-2 authors believe that placebo groups can still inform clinical decisions [12,49–53]. Dal-Re et al. also found PRECIS-2 insufficiently sensitive to differentiate between masked and open-label trials [51]. The placebo effects in "touch" interventions are considered part of MT treatments' therapeutic outcomes [54,55]. Although high-quality efficacy trials are required to clarify whether a manual therapy intervention is more than a placebo [56], including a placebo group in MT RCTs (although it has explanatory connotations) could aid in understanding the real effect of MT interventions that are challenging to achieve in an efficacy study under less realistic conditions. The ongoing PRECIS-3 update aims to assess control groups more effectively, highlighting the need for further research on the role of placebos in PT, which will be particularly interesting in MT [51–53,57].

Setting. Our results showed that many effectiveness trials were conducted in a single centre, limiting results' generalisability [49]. However, Zwarestein et al. argue that detailed reporting of the environment, health system, and patient types can still provide valuable clinical insights [50]. While multicentre trials are preferable for broader applicability, they are costly and challenging for MT researchers, who typically lack adequate public and private funding, a significant barrier to more extensive research.

Blinding. PRECIS-2 authors acknowledge that although blinding poses challenges, it remains beneficial in pragmatic studies by significantly enhancing decision-making data quality through the control of the placebo effect and subjectivity [50,57]. In MT, where mechanistic evidence and placebo-controlled studies are scarce, maintaining blinding in trials might help to preserve the rigor of RCTs.

Reporting and quality of effectiveness MT RCTs

Accurate trial reporting is essential for the applicability and quality of results [4,22]. Our review found that ETac were reported less comprehensively than PT and ETpl. It is crucial to include adequate details on trial settings, participants, and interventions to facilitate external implementation and validation [4]. In MT, biases are common due to challenges in blinding therapists and patients, especially when patient-reported outcomes are used. This supports a pragmatic approach but poses a high risk of bias, reducing internal validity.

Our findings indicate that the external validity of effectiveness trials is often compromised. A thorough justification for the trial's aim, rationale, and design can improve study quality and interpretation. Modifications to trial methodologies are essential to enhance their clinical applicability in the MT field. Clear and comprehensive reporting is vital for understanding trial quality, mitigating biases, and improving the applicability of findings for clinical practice.

STRENGTHS AND LIMITATIONS

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

CONCLUSIONS

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

ABBREVIATIONS

PRECIS-2: PRagmatic-Explanatory-Continuum-Indicatory-Summary
MT: Manual Therapy
RCT: Randomised Controlled Trial
PT: pragmatic Randomised Controlled Trial
ET: effectiveness Randomised Controlled Trial
ETac: effectiveness Randomised Controlled Trial using an active comparator
ETpl: effectiveness Randomised Controlled Trial using a placebo comparator
RoB: Risk of Bias

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

ACKNOWLEDGMENTS

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

BIBLIOGRAPHY

[1] Daniel Maddox C, Subialka JA, Young JL, Rhon DI. TITLE: Over Half of Clinical Trials of Mobilization and Manipulation for Patients with Low Back Pain May Have Limited Real-World Applicability. A Systematic Review pf 132 Clinical Trials n.d.
[2] Glasziou P, Matthews R, Boutron I, Chalmers I, Armitage P. The differences and overlaps between “explanatory” and “pragmatic” controlled trials: a historical perspective. J R Soc Med 2023:14107682312075 36. https://doi.org/10.1177/01410768231207536.
[3] Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. J Chronic Dis 1967;20:637–48. https://doi.org/10.1016/0021-9681(67)90041-0.
[4] Singal AG, Higgins PDR, Waljee AK. A primer on effectiveness and efficacy trials. Clin Transl Gastroenterol 2014;5:e45. https://doi.org/10.1038/ctg.2013.13.
[5] Janiaud P, Dal-Ré R, Ioannidis JPA. Assessment of Pragmatism in Recently Published Randomized Clinical Trials. JAMA Internal Medicine 2018;178:1278. https://doi.org/10.1001/jamainternmed.2018.3321.
[6] Dal-Ré R, Janiaud P, Ioannidis JPA. Real-world evidence: How pragmatic are randomized controlled trials labeled as pragmatic? BMC Med 2018;16:49. https://doi.org/10.1186/s12916-018-1038-2.
[7] Sacristán JA, Dilla T. Pragmatic trials revisited: applicability is about individualization. J Clin Epidemiol 2018;99:164–6. https://doi.org/10.1016/j.jclinepi.2018.02.003.
[8] Ford I, Norrie J. Pragmatic Trials. New England Journal of Medicine 2016;375:454–63. https://doi.org/10.1056/nejmra1510059.
[9] Fortney JC, Curran GM, Lyon AR, Check DK, Flum DR. Similarities and differences between Pragmatic Trials and Hybrid Effectiveness-Implementation Trials. J Gen Intern Med 2024;39:1735–43. https://doi.org/10.1007/s11606-024-08747-1.
[10] Nicholls SG, Zwarenstein M, Hey SP, Giraudeau B, Campbell MK, Taljaard M. The importance of decision intent within descriptions of pragmatic trials. J Clin Epidemiol 2020;125:30–7. https://doi.org/10.1016/j.jclinepi.2020.04.030.
[11] Treweek S, Zwarenstein M. Making trials matter: pragmatic and explanatory trials and the problem of applicability. Trials 2009;10:37. https://doi.org/10.1186/1745-6215-10-37.
[12] Thorpe KE, Zwarenstein M, Oxman AD, Treweek S, Furberg CD, Altman DG, et al. A pragmatic–explanatory continuum indicator summary (PRECIS): a tool to help trial designers. Journal of Clinical Epidemiology 2009;62:464–75. https://doi.org/10.1016/j.jclinepi.2008.12.011.
[13] Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, et al. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. BMJ 2008;337:a2390. https://doi.org/10.1136/bmj.a2390.
[14] Alvarez G, Solà I, Sitjà-Rabert M, Fort-Vanmeerhaeghe A, Gich I, Fernández C, et al. A methodological review revealed that reporting of trials in manual therapy has not improved over time. J Clin Epidemiol 2020;121:32–44. https://doi.org/10.1016/j.jclinepi.2020.01.006.
[15] 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. Journal of Clinical Epidemiology 2020;121:32–44. https://doi.org/10.1016/j.jclinepi.2020.01.006.

[16] Riley SP, Swanson BT, Sawyer SF, Brismée J-M. Is research quality in orthopedic manual therapy trials stagnating? Reflections and pathways for improving research quality and advance our profession. J Man Manip Ther 2016;24:239–40. https://doi.org/10.1080/10669817.2016.1253561.
[17] Gonzalez GZ, Moseley AM, Maher CG, Nascimento DP, Costa L da CM, Costa LO. Methodologic Quality and Statistical Reporting of Physical Therapy Randomized Controlled Trials Relevant to Musculoskeletal Conditions. Arch Phys Med Rehabil 2018;99:129–36. https://doi.org/10.1016/j.apmr.2017.08.485.
[18] Koes BW. How to evaluate manual therapy: value and pitfalls of randomized clinical trials. Man Ther 2004;9:183–4. https://doi.org/10.1016/j.math.2004.04.002.
[19] 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. https://doi.org/10.1016/j.jclinepi.2020.06.025.
[20] Clar C, Tsertsvadze A, Court R, Hundt GL, Clarke A, Sutcliffe P. Clinical effectiveness of manual therapy for the management of musculoskeletal and non-musculoskeletal conditions: systematic review and update of UK evidence report. Chiropr Man Therap 2014;22:12. https://doi.org/10.1186/2045-709X-22-12.
[21] Riley SP, Swanson B, Brismée J-M, Sawyer SF. A systematic review of orthopaedic manual therapy randomized clinical trials quality. Journal of Manual & Manipulative Therapy 2016;24:241–52. https://doi.org/10.1080/10669817.2015.1119372.
[22] Innocenti T, Giagio S, Salvioli S, Feller D, Minnucci S, Brindisino F, et al. Completeness of Reporting Is Suboptimal in Randomized Controlled Trials Published in Rehabilitation Journals, With Trials With Low Risk of Bias Displaying Better Reporting: A Meta-research Study. Archives of Physical Medicine and Rehabilitation 2022. https://doi.org/10.1016/j.apmr.2022.01.156.
[23] Fregni F, Imamura M, Chien HF, Lew HL, Boggio P, Kaptchuk TJ, et al. Challenges and recommendations for placebo controls in randomized trials in physical and rehabilitation medicine: a report of the international placebo symposium working group. Am J Phys Med Rehabil 2010;89:160–72. https://doi.org/10.1097/PHM.0b013e3181bc0bbd.
[24] Simoni G, Bozzolan M, Bonnini S, Grassi A, Zucchini A, Mazzanti C, et al. Effectiveness of standard cervical physiotherapy plus diaphragm manual therapy on pain in patients with chronic neck pain: A randomized controlled trial. J Bodyw Mov Ther 2021;26:481–91. https://doi.org/10.1016/j.jbmt.2020.12.032.
[25] Scafoglieri A, Van den Broeck J, Willems S, Tamminga R, van der Hoeven H, Engelsma Y, et al. Effectiveness of local exercise therapy versus spinal manual therapy in patients with patellofemoral pain syndrome: medium term follow-up results of a randomized controlled trial. BMC Musculoskelet Disord 2021;22:446. https://doi.org/10.1186/s12891-021-04310-9.
[26] Muñoz-Gómez E, Inglés M, Serra-Añó P, Espí-López GV. Effectiveness of a manual therapy protocol based on articulatory techniques in migraine patients. A randomized controlled trial. Musculoskelet Sci Pract 2021;54:102386. https://doi.org/10.1016/j.msksp.2021.102386.
[27] Izaola-Azkona L, Vicenzino B, Olabarrieta-Eguia I, Saez M, Lascurain-Aguirrebeña I. Effectiveness of Mobilization of the Talus and Distal Fibula in the Management of

Acute Lateral Ankle Sprain. Phys Ther 2021;101. https://doi.org/10.1093/ptj/pzab111.
[28] Gamerman V, Cai T, Elsäßer A. Pragmatic randomized clinical trials: best practices and statistical guidance. Health Serv Outcomes Res Methodol 2019;19:23–35. https://doi.org/10.1007/s10742-018-0192-5.
[29] Wieland LS, Berman BM, Altman DG, Barth J, Bouter LM, D'Adamo CR, et al. Rating of Included Trials on the Efficacy-Effectiveness Spectrum: development of a new tool for systematic reviews. J Clin Epidemiol 2017;84:95–104. https://doi.org/10.1016/j.jclinepi.2017.01.010.
[30] Deaton A, Cartwright N. Understanding and misunderstanding randomized controlled trials. Soc Sci Med 2018;210:2–21. https://doi.org/10.1016/j.socscimed.2017.12.005.
[31] Porzsolt F, Wiedemann F, Phlippen M, Weiss C, Weiss M, Schmaling K, et al. The terminology conflict on efficacy and effectiveness in healthcare. J Comp Eff Res 2020;9:1171–8. https://doi.org/10.2217/cer-2020-0149.
[32] Hohenschurz-Schmidt D, Kleykamp BA, Draper-Rodi J, Vollert J, Chan J, Ferguson M, et al. Pragmatic trials of pain therapies: a systematic review of methods. Pain 2021.
[33] Palakshappa JA, Gibbs KW, Lannan MT, Cranford AR, Taylor SP. Systematic Review of the "Pragmatism" of Pragmatic Critical Care Trials. Critical Care Explorations 2022;4:e0738. https://doi.org/10.1097/CCE.0000000000000738.
[34] Roura S, Alvarez G, Hohenschurz-Schmidt D, Solà I, Núñez-Cortés R, Bracchiglione J, et al. Lack of pragmatic attitude of self-labelled pragmatic trials on manual therapy: a methodological review. BMC Med Res Methodol 2024;24. https://doi.org/10.1186/s12874-024-02393-1.
[35] Fritz JM, Cleland J. Effectiveness versus efficacy: more than a debate over language. J Orthop Sports Phys Ther 2003;33:163–5. https://doi.org/10.2519/jospt.2003.33.4.163.
[36] Witt CM. Efficacy, effectiveness, pragmatic trials--guidance on terminology and the advantages of pragmatic trials. Forsch Komplementmed 2009;16:292–4. https://doi.org/10.1159/000234904.
[37] Gordon KS, Peduzzi P, Kerns RD. Designing Trials with Purpose: Pragmatic Clinical Trials of Nonpharmacological Approaches for Pain Management. Pain Med 2020;21:S7–12. https://doi.org/10.1093/pm/pnaa347.
[38] Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane Handbook Systematic Reviews Interventions version 6. Cochrane: 2024.
[39] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Rev Esp Cardiol 2021;74:790–9. https://doi.org/10.1016/j.rec.2021.07.010.
[40] Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016;5:210. https://doi.org/10.1186/s13643-016-0384-4.
[41] Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. BMJ 2015;350:h2147. https://doi.org/10.1136/bmj.h2147.
[42] Devos F, Foissac F, Bouazza N, Ancel P-Y, Tréluyer J-M, Chappuy H. Study characteristics impacted the pragmatism of randomized controlled trial published

in nursing: a meta-epidemiological study. J Clin Epidemiol 2019;116:18–25. https://doi.org/10.1016/j.jclinepi.2019.07.017.
[43] Elder WG, Munk N. Using the Pragmatic-Explanatory Continuum Indicator Summary (PRECIS) Model in Clinical Research: Application to Refine a Practice-based Research Network (PBRN) Study. The Journal of the American Board of Family Medicine 2014;27:846–54. https://doi.org/10.3122/jabfm.2014.06.140042.
[44] Rosas LG, Lv N, Azar K, Xiao L, Yank V, Ma J. Applying the Pragmatic–Explanatory Continuum Indicator Summary Model in a Primary Care–Based Lifestyle Intervention Trial. Am J Prev Med 2015;49:S208–14. https://doi.org/10.1016/j.amepre.2015.05.011.
[45] Roura, Alvarez, Hohenschurz-Schmidt D, Solà, Núñez-Cortés R, Bracchiglione, et al. A call for improving the Internal validity and the reporting of manual therapy trials self-labelled as pragmatic: A methodological review. Int J Osteopath Med 2025:100754. https://doi.org/10.1016/j.ijosm.2025.100754.
[46] Bishop MD, Torres-Cueco R, Gay CW, Lluch-Girbés E, Beneciuk JM, Bialosky JE. What effect can manual therapy have on a patient's pain experience? Pain Manag 2015;5:455–64. https://doi.org/10.2217/pmt.15.39.
[47] Lin I, Wiles L, Waller R, Goucke R, Nagree Y, Gibberd M, et al. What does best practice care for musculoskeletal pain look like? Eleven consistent recommendations from high-quality clinical practice guidelines: systematic review. Br J Sports Med 2020;54:79–86. https://doi.org/10.1136/bjsports-2018-099878.
[48] Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ 2014;348:g1687. https://doi.org/10.1136/bmj.g1687.
[49] Dal-Ré R, de Boer A, James SK. The design can limit PRECIS-2 retrospective assessment of the clinical trial explanatory/pragmatic features. J Clin Epidemiol 2020;126:193–201. https://doi.org/10.1016/j.jclinepi.2020.03.027.
[50] Zwarenstein M, Thorpe K, Treweek S, Loudon K. PRECIS-2 for retrospective assessment of RCTs in systematic reviews. J Clin Epidemiol 2020;126:202–6. https://doi.org/10.1016/j.jclinepi.2020.06.023.
[51] Dal-Ré R. The PRECIS-2 tool seems not to be useful to discriminate the degree of pragmatism of medicine masked trials from that of open-label trials. Eur J Clin Pharmacol 2021;77:539–46. https://doi.org/10.1007/s00228-020-03030-8.
[52] Dal-Ré R. Pragmatic trials, blinding, placebos, and the usefulness of the PRECIS-2 tool. Eur J Clin Pharmacol 2021;77:1071–2. https://doi.org/10.1007/s00228-020-03079-5.
[53] Zwarenstein M, Howie A. Blinding, pragmatism, and the PRECIS-2 tool for designing and assessing randomized trials. Eur J Clin Pharmacol 2021;77:1069–70. https://doi.org/10.1007/s00228-020-03078-6.
[54] Rossettini G, Carlino E, Testa M. Clinical relevance of contextual factors as triggers of placebo and nocebo effects in musculoskeletal pain. BMC Musculoskelet Disord 2018;19:27. https://doi.org/10.1186/s12891-018-1943-8.
[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. https://doi.org/10.2519/jospt.2017.0604.
[56] Hohenschurz-Schmidt D, Liem T. Placebo effects in osteopathy and other manual therapies – what they are and why they matter to clinical practice, education, and

research. Int J Osteopath Med 2025:100762.
https://doi.org/10.1016/j.ijosm.2025.100762.
[57] Willis A, Shiely F, Treweek S, Taljaard M, Loudon K, Howie A, et al. Comments, suggestions and criticisms of the PRECIS-2 design tool: A citation analysis. J Clin Epidemiol 2024:111534. https://doi.org/10.1016/j.jclinepi.2024.111534.