Musculoskeletal Science and Practice 81 (2026) 103480

Contents lists available at ScienceDirect

Musculoskeletal Science and Practice

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

Review article

Latent classes of trial reporting and publication practices in spinal manipulation research: a meta-epidemiological study

Casper Nim[a,b,c,*], Michelle Frederiksen[c], Sasha Aspinall[d], Aron Downie[e], Martha Funabashi[c,f,g,h], Steen Harsted[c,i], Hazel Jenkins[e], David McNaughton[j], Luana Nyirö[k], Eric J. Roseen[l], James J. Young[c,m], Liz Dennett[n], Stephen M. Perle[o], Chad Cook[p,q,r], Carsten Juhl[c,s], Jan Hartvigsen[c,t]

[a] Medical Research Unit, Spine Centre of Southern Denmark, Lillebaelt Hospital, University Hospital of Southern Denmark, Kolding, Denmark
[b] Department of Regional Health Research, University of Southern Denmark, Odense, Denmark
[c] Center for Muscle and Joint Health, Department of Sports Science and Clinical Biomechanics, University of Southern Denmark, Odense, Denmark
[d] School of Allied Health, College of Health and Education, Murdoch University, Perth, Australia
[e] Department of Chiropractic, Macquarie University, Sydney, Australia
[f] Division of Research and Innovation, Canadian Memorial Chiropractic College, Toronto, Canada
[g] Department of Chiropractic, Université du Québec à Trois-Rivières, Trois-Rivières, Canada
[h] Research Center, Parker University, Dallas, USA
[i] SDU Health Informatics and Technology, The Maersk Mc-Kinney Moller Institute, Odense, Denmark
[j] College of Health, Medical and Applied Sciences, CQUniversity, Brisbane, Australia
[k] Department of Chiropractic Medicine, Balgrist University Hospital, University of Zurich, Zurich, Switzerland
[l] Section of General Internal Medicine, Department of Medicine, Boston University, Chobanian & Avedisian School of Medicine and Boston Medical Center, Boston, MA, USA
[m] Schroeder Arthritis Institute, Krembil Research Institute, University Health Network, Toronto, Canada
[n] Geoffrey and Robyn Sperber Health Sciences Library, University of Alberta, Edmonton, Canada
[o] Big Data Interrogation Group, Health Sciences University, Bournemouth, UK
[p] Department of Orthopaedics, Duke University, Durham, NC, USA
[q] Duke Clinical Research Institute, Duke University, Durham, North Carolina, USA
[r] Department of Population Health Sciences, Duke University, Durham, North Carolina, USA
[s] Department of Physiotherapy and Occupational Therapy, Copenhagen University Hospital, Herlev and Gentofte, Denmark
[t] Chiropractic Knowledge Hub, Odense, Denmark

ARTICLE INFO

Keywords:
Research integrity
Research waste
Latent class analysis
Spinal manipulation

ABSTRACT

Background: Reliable reporting and publication practices are essential for trustworthy evidence synthesis and clinical decision-making.
Objective: We aimed to identify latent classes of randomized controlled trials (RCTs) evaluating spinal manipulative therapy (SMT) based on trial reporting and publication practices, and to examine whether these classes influenced treatment effects.
Design: Meta-epidemiological study.
Method: Trials were evaluated on whether they met criteria for trial reporting and publication practices across six domains. Latent class analysis was used to identify trial subgroups. Random-effects meta-regression models assessed whether class membership predicted pooled estimates of treatment effects for pain and disability.
Results: We included 239 RCTs and identified four classes: *Dated* (23 %), older trials (mostly pre-2010) with consistently low proportions of criteria met; *Non-contributing* (30 %), newer trials that inconsistently met the criteria, had small samples, and short follow-ups; *SMT-focused* (15 %), which reported SMT details and fidelity more consistently but otherwise resembled the *Non-contributing* class; and *Pragmatic* (33 %), consisting of larger trials, meeting most criteria, but often underreported SMT-specific and fidelity details. Reporting practices had

* Corresponding author. Medical Research Unit, Spine Centre of Southern Denmark, Lillebaelt Hospital, University Hospital of Southern Denmark, Kolding, Denmark.
E-mail address: casper.nim@rsyd.dk (C. Nim).

https://doi.org/10.1016/j.msksp.2025.103480
Received 10 October 2025; Received in revised form 23 December 2025; Accepted 27 December 2025
Available online 27 December 2025
2468-7812/

larger impact on class membership than publication practices. Despite differences class membership was not associated with treatment effect estimates and explained minimal outcome variability (R^2 ~1 %).
Conclusions: Although trial reporting and publication practices varied substantially across SMT trials, these differences were not associated with differences in treatment effects. The widespread failure to meet key criteria raises concerns about the interpretability and credibility of the SMT evidence base. To strengthen transparency and scientific value, future trials should adhere more rigorously to reporting guidelines.

1. Introduction

Spinal pain is prevalent and remains a leading cause of disability worldwide (Ferreira et al., 2023; Wu et al., 2024). Evidence-based clinical guidelines endorse various treatment strategies, frequently including spinal manipulative therapy (SMT) (Corp et al., 2021; World Health Organization, 2023). SMT has been studied extensively, as evidenced by the publication of over 130 systematic reviews on its effectiveness (Eybye et al., 2022). Yet, the methodological transparency of randomized controlled trials (RCTs) in this domain remains limited. Critical intervention details, such as manual therapy techniques used, clinical reasoning behind SMT application, and co-interventions, are frequently underreported (Groeneweg et al., 2017; Ruzich et al., 2024). Although guidelines like CONSORT aim to standardize and improve reporting (Moher et al., 2010), adherence has been inconsistent (Alvarez et al., 2020). Incomplete reporting impairs replication, hinders risk of bias (RoB) assessment, and can bias meta-analytic findings, potentially leading to misguided clinical recommendations.

While RoB is routinely assessed in systematic reviews (Higgins et al., 2022), other factors, such as pre-registration, documentation quality, fidelity monitoring, and publication practices, receive less attention, despite their relevance to transparency and trust. Hayden et al. summarized the results of RCTs evaluating the effectiveness of exercise therapy for chronic low back pain and demonstrated that poor reporting and publication practices were associated with larger effect sizes for both pain and disability outcomes (Hayden et al., 2021). Nonetheless, when employing a data-driven approach that included over 40 trial characteristics, trials classified as having “Fundamental Deficiencies” only slightly overestimated treatment effects for disability, and not for pain outcomes (Hayden et al., 2025).

SMT and exercise therapy are often evaluated within the same broad domain of conservative, non-pharmacological interventions (Foster et al., 2018), however, several conceptual and methodological features, that may affect latent patterns, differ between trials dealing with SMT and exercise. SMT is a passive, therapist-delivered intervention, whereas exercise is an active approach, which influence the types of comparators typically used in each field and the contextual effects of the two interventions (Bejarano et al., 2023; Snow et al., 2019). These differences create uncertainty about whether SMT trials will cluster into latent classes like those observed in exercise therapy trials (Hayden et al., 2025), or whether the relationship between class membership and treatment effects will diverge. Identifying latent classes specifically within SMT trials therefore adds value beyond the previous study by allowing us to test whether the latent class structures identified in exercise research replicate in this conceptually distinct intervention domain. If the patterns differ, this suggests that methodological challenges and reporting behaviors are intervention-specific; if they are similar, it strengthens the generalizability of latent-class approaches across conservative, non-pharmacological spinal pain research. In either case, examining SMT trials separately provides insight into whether reporting and publication characteristics influence effect estimates. Finally, persistent deficiencies in these areas compromise interpretability and undermine confidence in the evidence base, and the growing number of journals with limited editorial oversight, often termed “predatory” (Sénéquier et al., 2025; Candal-Pedreira et al., 2022; Soriano and Ruano-Ravina, 2025), further highlight the importance of evaluating the integrity of SMT trial reporting.

We aimed to address these issues in relation to RCTs involving SMT in the treatment of spinal pain. We a) identified latent classes based on trial reporting and publication practices, and b) assessed whether class-membership influenced estimates of SMT treatment effects for pain and disability outcomes.

2. Method

2.1. Design

We conducted a meta-epidemiological analysis based on a protocol for a network meta-analysis on SMT procedures (Nim et al., 2023, 2025). The present study followed an updated protocol registered in the Open Science Framework (OSF; Protocol 2, https://osf.io/t4dpe). An original protocol (Protocol 1, https://osf.io/se8rf) was initially developed following the meta-epidemiological framework described by Hayden et al. (2021), which focused on a more limited set of reporting and publication characteristics (Higgins et al., 2022). During the preparatory phase of the current study, and prior to any outcome analyses, Protocol 1 was substantively revised to align with the approach introduced by Hayden et al. (2025).

Protocol 2 reflects a conceptual and methodological shift from evaluating a small number of predefined summary indicators to modeling a substantially expanded set of individual trial-level items related to trial planning, conduct, reporting, and publication practices using a data-driven classification approach. Specifically, the updated protocol expanded both the number and granularity of assessed items, with additional emphasis on SMT-specific intervention characteristics and fidelity domains. These changes were made to better capture heterogeneity in SMT trial design and reporting and to enable identification of latent methodological patterns that could not be adequately represented using Protocol 1.

2.2. Eligibility criteria

We included RCTs involving adults with spinal pain (in any region) treated with SMT. We included SMT trials in which high-velocity, low-amplitude SMT was delivered either as a standalone technique or in combination with other manual therapy interventions such as low-velocity spinal mobilization. We included trials using any comparator that allowed the SMT effect to be isolated. Eligible comparators included:

- Inactive controls, such as sham SMT, placebo, detuned devices, minimal treatment, or no-treatment controls
- Active interventions, such as exercise therapy, manual therapy, other non-SMT treatments
- Other SMT variants, where different SMT procedures were compared within the same study.

Studies written in English, Spanish, Portuguese, German, or Scandinavian languages were included.

2.3. Trial selection and data extraction

Systematic reviews were identified via PubMed and Epistemonikos (through Feb 25, 2022) (Eybye et al., 2022). A subsequent systematic search was performed from January 1, 2018, to September 12, 2023 (updated December 22nd, 2024), across MEDLINE, EMBASE, CENTRAL, PEDro, and Index to Chiropractic Literature. Two reviewers independently screened studies, extracted data, and assessed risk of bias, with CN resolving conflicts. Most extracted data are detailed in the original protocol (Nim et al., 2023), with additional data extracted for this analysis.

2.4. Definitions of trial reporting, conduct and publication factors

Two authors independently assessed *trial reporting and publication practice domains:* Trial planning (5 items), Trial conduct (6 items), Trial reporting – design (3 items), Trial reporting - participants (8 items), Trial reporting - treatment (16 items), Trial reporting – outcomes (4 items), and Publication practices (5 items), where a score of 1 was assigned for each criteria that was met (e.g., reported or low risk of bias) (Supplementary file 1).

Some definitions were modified from Hayden et al. (2025). After extensive discussion within the research team regarding publication practices, including information specialists/research librarians, we concluded that a clear definition of "predatory publishing" lacked reliable consensus. Therefore, for publication practices, we assessed time in peer-review, retraction status (retraction status was verified on www.retractiondatabase.org on March 15, 2025), and whether the journal was a member of a "verification-site" (i.e., COPE, DOAJ, and JCR). Another notable change was the use of Cochrane RoB Tool 2 (RoB2) instead of version 1. Furthermore, to avoid circular reasoning, we excluded pain and disability outcomes from our modeling. Finally, modifications were made to reflect the SMT context (Supplementary file 1).

2.5. Statistical analysis

2.5.1. Latent class analysis

We used Latent Class Analysis (LCA) to identify classes of trials based on binary indicators across reporting and publication domains (1 = meeting the criteria, 0 = not meeting criteria). Models with 1–7 classes were fitted using expectation-maximization with multiple random starts to ensure convergence (Soriano and Ruano-Ravina, 2025). The optimal number of classes was selected using AIC, BIC, entropy, average posterior probabilities, proportions with low posterior probabilities, and class sizes (Linzer and Lewis, 2011). *A-priori,* we expected four classes based on Hayden et al. (2025), however, as there were differences across variables, we opted for an exploratory modeling approach. We labeled the classes based on consensus among the research team guided by the following:

- The conditional proportions of classes for each publication year via loess plots with years binned in "before 2000", and then 5-year increments.
- The contribution of each item within the latent classes plotted as the distribution of each reporting item within that class.
- The class distribution for global region, longest follow-up, number of sessions, number of participants, and comparator

Class-specific probabilities for each item were derived from item means within each class. Differences between classes were assessed using chi^2 tests ($\alpha = 0.05$, no correction for multiplicity due to exploratory aim). Heat maps visualized the scoring patterns.

2.5.2. Effect estimates

Pain outcomes were rescaled to a 0–100 scale and analyzed as mean differences (MD) (Wewege et al., 2022), while disability outcomes were analyzed as standardized mean differences (SMD) (Nim et al., 2025). We focused on short-term (post-treatment) and long-term (closest to 12-month) outcomes. Comparators included inactive controls (no/sham/minimal treatment) and active interventions. In trials with multiple SMT or control arms, groups were merged to yield one contrast per study using an assumed correlation of 0.6 for change scores (Higgins et al., 2022).

2.5.3. Associations between classes and SMT effect size

We used stratified meta-analyses to evaluate whether latent class membership influenced SMT treatment effects for pain and disability at short-term and long-term follow-up. Random-effects meta-regression models were fitted with trial class as a categorical moderator, using the least "well-reported" class as the reference. Between-class differences in treatment effects were assessed using class-specific coefficients and Q-tests for heterogeneity. We evaluated model fit and residual heterogeneity using I^2, τ^2. We interpreted findings using thresholds for clinical relevance: MD $\geq$ 20/SMD$\geq$0.5 for inactive controls, and MD $\geq$ 10/SMD$\geq$0.3 for active comparisons (Nim et al., 2025). Sensitivity analyses excluded trials with posterior probability $<$0.8 (to examine potential misclassification bias) and stratified results by country income group.

All analyses were conducted in R (R Core Team, 2023) using *Tidyverse* (Wickham et al., 2019). LCA was conducted using *poLCA* (Linzer and Lewis, 2011). Meta-analyses were performed using the *metafor* package (Viechtbauer, 2010).

2.6. Protocol amendments

We made minor changes to improve Interpretability of Protocol 2. We included additional characteristics related to adverse events. For conceptual reasons we separated the *Trial reporting – treatment* domain into i) a general treatment-reporting domain, and ii) an SMT-specific reporting domain capturing technical details unique to SMT, such as positioning, targeting, thrust characteristics, and fidelity procedures.

3. Results

The SMT database consisted of 244 studies; we included 239 studies for analysis, published in 90 journals. Five papers were excluded as they did not meet the language inclusion criterion.

3.1. Latent class analysis

Several variables were excluded from the LCA due to limited variability or inconsistent reporting: country of conduct, population source (correlated with study setting), SMT intervention region, SMT thrust approach and retraction status (no studies had been retracted).

We selected a four-class solution based on *a-priori* expectations and statistical fit indices. This model yielded high average posterior probabilities (>0.95), had the lowest BIC, and maintained acceptable class size (smallest class = 15 %) (Supplementary file 2). Only nine studies had posterior probabilities below 0.8, and three fell below 0.6. Classes were labeled using temporal patterns (Fig. 1), class-level response distributions (Fig. 2), and study characteristics (Table 1). Most domains showed variation in whether trials met criteria across classes. Thus, labeling was directed towards temporal distribution and study characteristics. The four classes were: *Dated* (n = 54, 23 %), *Non-contributing* (n = 71, 30 %), *SMT-focused* (n = 36, 15 %), and *Pragmatic* (n = 78, 33 %) (Supplementary file 3).

Temporal trends showed that *Dated* studies were concentrated before 2010. The *SMT-focused* class remained relatively stable but small across decades. In contrast, *Non-contributing* increased sharply since 2000, comprising nearly 50 % of recent publications. *Pragmatic* peaked in the early 2000s (Fig. 1).

The *Dated* class showed the lowest overall proportions of trials meeting criteria. The *SMT-focused* class most often met criteria for SMT-specific items such as tailoring, positioning, session details, and treatment fidelity variables. The *Non-contributing* class showed variation across domains but was notably poor in meeting criteria for adverse events. The *Pragmatic* class also showed variation but generally met

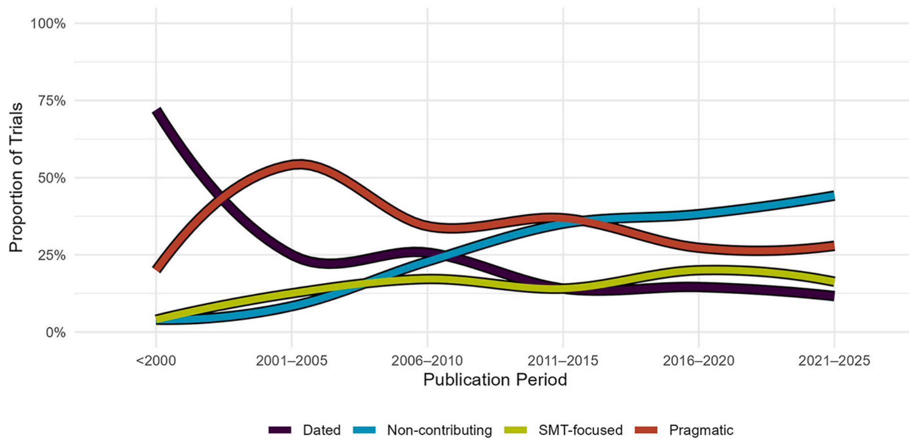

Fig. 1. Temporal distribution of included trials by latent classes.

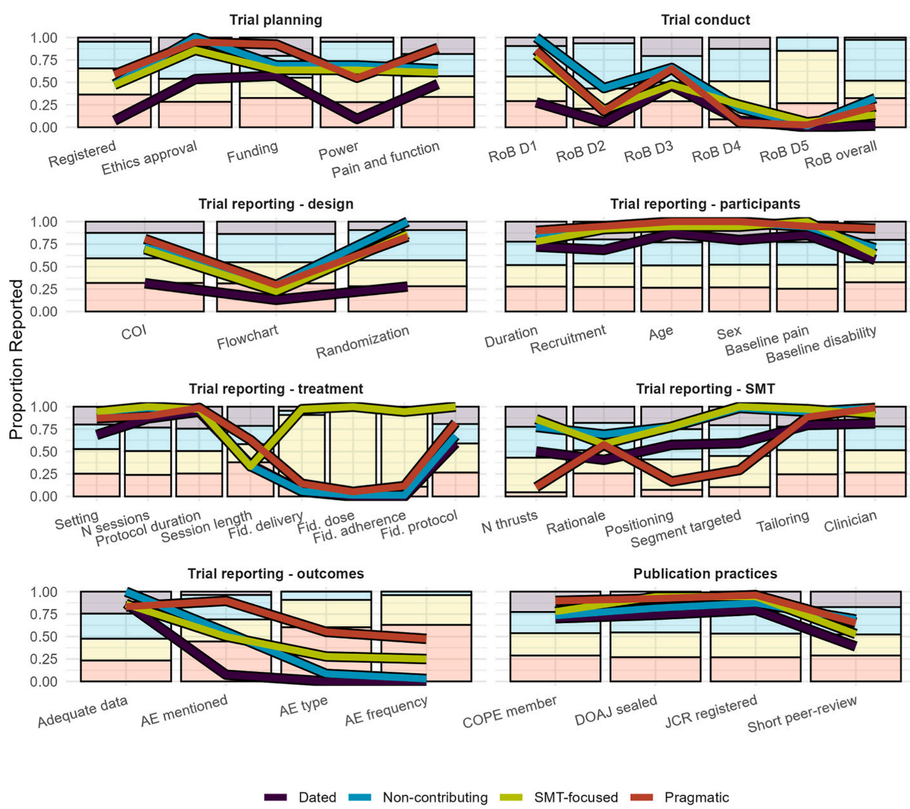

Fig. 2. Mean proportions of trials meeting criteria across trial and publication domains by latent class
Lines indicate the mean proportion of meeting the criteria (scoring "1") and the shaded area indicates the distribution across classes within each characteristic. RoB = Risk of bias, RoB D1 = Bias arising from the randomization process, RoB D2 = Bias due to deviations from intended interventions, RoB D3 = Bias due to missing outcome data, RoB D4 = Bias in measurement of the outcome, RoB D5 = Bias in selection of the reported result, COI = Conflict of interest, Fid. = Fidelity procedure, AE = Adverse event, COPE = Committee on Publication Ethics, DOAJ = Directory of Open Access Journals, JCR = Journal Citation Reports.

criteria more often across domains, except for SMT-specific reporting. Publication practices showed little variation between classes (Fig. 2).

Most trials were conducted in North America (36 %) and Europe (32 %), though class distribution varied by region. The proportion of trials classified as *Dated* or *Non-contributing* was highest in Asia (76 %, n = 25) and South America (68 %, n = 13), followed by Europe (55 %, n = 42), Africa (50 %, n = 4), North America (44 %, n = 38), and Oceania (19 %, n = 3). Sample sizes were generally small (median 24–27 participants), except in the *Pragmatic* class (median of 45). This class also stood out for delivering more sessions and longer follow-up. In contrast, other classes tended to assess outcomes immediately after treatment and delivered ≤4 sessions. Comparator types were evenly distributed across classes, though *SMT-focused* trials more often lacked an active comparator group (Table 1).

3.2. Conditional probability of reporting for each item in each class

The *Dated* class consistently had the lowest probabilities of meeting criteria across nearly all domains, particularly for trial planning and trial conduct. The *Pragmatic* class generally demonstrated the highest probabilities of meeting criteria, especially in participant characteristics and treatment setting variables.

Table 1
Descriptive characteristics of trials by latent class.

Study characteristics	Overall N = 239	Dated N = 54	Non-contributing N = 71	SMT-focused N = 36	Pragmatic N = 78
Global region					
Africa	8 (3.3 %)	2 (3.7 %)	2 (2.8 %)	3 (8.3 %)	1 (1.3 %)
Asia	33 (14 %)	13 (24 %)	12 (17 %)	2 (5.6 %)	6 (7.7 %)
Europe	76 (32 %)	15 (28 %)	27 (38 %)	15 (42 %)	19 (24 %)
North America	87 (36 %)	18 (33 %)	20 (28 %)	14 (39 %)	35 (45 %)
Oceania	16 (6.7 %)	2 (3.7 %)	1 (1.4 %)	0 (0 %)	13 (17 %)
South America	19 (7.9 %)	4 (7.4 %)	9 (13 %)	2 (5.6 %)	4 (5.1 %)
Number of participants, Median (Q1-Q3)	30 (19, 51)	24 (15, 43)	27 (18, 41)	27 (17, 38)	45 (22, 73)
Comparator					
Active	141 (59 %)	32 (59 %)	40 (56 %)	13 (36 %)	56 (72 %)
Inactive	63 (26 %)	14 (26 %)	20 (28 %)	10 (28 %)	19 (24 %)
Both active and inative	11 (4.6 %)	3 (5.6 %)	2 (2.8 %)	4 (11 %)	2 (2.6 %)
Only SMT groups	24 (10 %)	5 (9.3 %)	9 (13 %)	9 (25 %)	1 (1.3 %)
Longest follow-up					
>6 weeks of follow-up	82 (34 %)	8 (15 %)	23 (32 %)	9 (25 %)	42 (54 %)
≤6 weeks of follow-up	33 (14 %)	7 (13 %)	15 (21 %)	5 (14 %)	6 (7.7 %)
Only immediate follow-up	124 (52 %)	39 (72 %)	33 (46 %)	22 (61 %)	30 (38 %)
Sessions					
≤4 sessions	91 (38 %)	20 (37 %)	41 (58 %)	21 (58 %)	9 (12 %)
4–6 sessions	50 (21 %)	11 (20 %)	12 (17 %)	7 (19 %)	20 (26 %)
>6 sessions	82 (34 %)	16 (30 %)	17 (24 %)	8 (22 %)	41 (53 %)
Not reported	16 (6.7 %)	7 (13 %)	1 (1.4 %)	0 (0 %)	8 (10 %)

Treatment reporting varied across classes. General items such as number of sessions and setting were often well reported, but treatment fidelity was rarely met outside the *SMT-focused* class, which demonstrated near-complete reporting across all fidelity items. This class also provided detailed reporting of SMT-specific features including patient positioning, targeting, and tailoring. By contrast, the *Pragmatic* class, despite meeting more criteria overall, reported fewer SMT-specific and fidelity details. The *Non-contributing* class varied greatly and particularly inconsistent in reporting of adverse events. Mentions of adverse events were frequent in the *Pragmatic* class (89 %), yet reporting of frequency or type remained low across all classes (Fig. 3).

3.3. Summary of the four classes

The *Dated* class consisted primarily of trials published before 2010 and showed consistently low probabilities of meeting criteria across most domains, including registration, sample size planning, risk of bias, fidelity reporting, and adverse events.

The *Non-contributing* class increased sharply after 2000 and represented nearly half of recent publications; this class also showed low and variable reporting across multiple domains, particularly adverse events, and included mostly small trials with limited follow-up.

The *SMT-focused* class was small but stable over time and demonstrated the most complete reporting of SMT-specific and fidelity-related items, and often compared different SMT procedures, although sample sizes and follow-up durations were generally small.

The *Pragmatic* class peaked in the early 2000s and comprised the largest trials, with more sessions and longer follow-up, and showed higher probabilities of meeting general reporting criteria, although SMT-specific details and fidelity items were less frequently reported.

3.4. Effect estimates

For pain outcomes, the estimated MDs consistently favored SMT (i.e., negative values), but the magnitude varied across classes. Despite this, at the end of treatment, there was no statistically significant class-differences for either SMT vs. Active ($p = 0.32$; $I^2 = 99$ %) or SMT vs. Inactive *(sham, placebo, no-treatment)* ($p = 0.40$; $I^2 = 90$ %). Similarly, at ~12-month follow-up, class-differences remained non-significant for both comparisons (Fig. 4). For disability outcomes, SMDs also tended to favor SMT, with the largest effects observed in *Non-contributing* and *SMT-focused* when compared to Active comparisons. However, none of the class-differences was statistically significant at either time point (Fig. 5). Across all comparisons, the between-study heterogeneity remained high, with I^2 values ranging from 89 % to 99 %, indicating considerable variability not attributable to sampling error. The explored classes explained very little of the heterogeneity (R^2 between 0 and 0.03) (Supplementary file 4).

3.5. Sensitivity analysis

Excluding the nine trials with a posterior probability below 0.8 did not change meta-analysis results (Supplementary File 5). We also explored an interaction between class and country income group (World Bank Country and Lending Groups) but could not proceed due to insufficient variation – specifically, the *SMT-focused* class included only high-income country trials, rendering interaction terms non-identifiable.

4. Discussion

This meta-epidemiological study identified four classes of reporting and publication characteristics among SMT trials. Although only a few domains sharply differentiated the classes, they also diverged on several trial-level features, supporting the presence of meaningful methodological patterns within the literature. The *Dated* class reflected historical reporting norms prior to the widespread adoption of modern standards. The *Non-contributing* and *SMT-focused* classes continued to show inconsistent reporting despite being composed largely of more recent trials. Notably, these two classes together represented 45 % of all included trials, and the *Non-contributing* class accounted for nearly half of all trials published between 2021 and 2025. The main distinction between these two classes was that the SMT-focused class consistently provided detailed SMT-specific and fidelity information. The *Pragmatic* class, which included larger trials such as the UK BEAM (UK BEAM Trial Team, 2004), UCLA (Hurwitz et al., 2002), and the Bronfort trials (Bronfort et al., 1996, 2001, 2012) tended to reflect real-world clinical practice but still demonstrated limited reporting of SMT-specific and fidelity elements.

All classes did not meet the RoB criteria except for randomization, which was expected given that bias is a common finding in systematic reviews of SMT (Rubinstein et al., 2019; Minnucci et al., 2023). This includes the primary analysis of this dataset (Nim et al., 2025), and the challenges in blinding manual therapy interventions (domain 4 of RoB2) (Muñoz Laguna et al., 2025). Reporting on fidelity was a common concern in three classes. This may reflect that attention to and assessment of fidelity is limited in healthcare researchers (McGee et al., 2018).

Despite meaningful class distinctions, class membership explained little to no variability in effect estimates for pain or disability outcomes. Heterogeneity remained high and none of the classes was associated

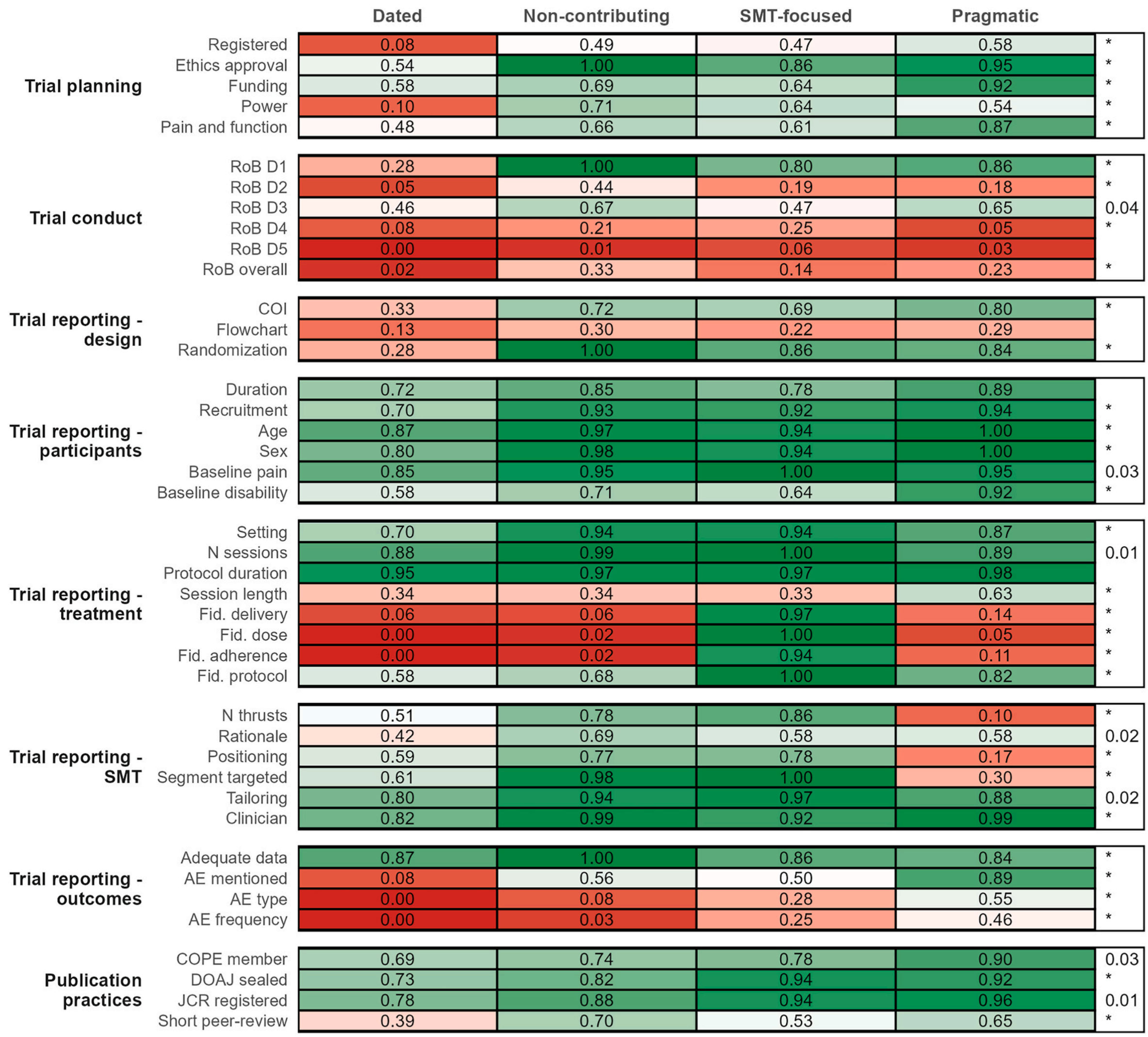

Fig. 3. Conditional probabilities across trial and publication domains reporting by latent class
Dark red indicates low probability of meeting the criteria (near 0), and dark green indicates high probability (near 1). Values on the far right indicate statistically significant differences between classes using Chi^2 tests, * indicates a p-value <0.01, AE = adverse event, RoB = Risk of bias, RoB D1 = Bias arising from the randomization process, RoB D2 = Bias due to deviations from intended interventions, RoB D3 = Bias due to missing outcome data, RoB D4 = Bias in measurement of the outcome, RoB D5 = Bias in selection of the reported result, COI = conflict of interest, DOAJ = Directory of Open Access Journals, COPE = Committee on Publication Ethics, JCR = Journal Citation Reports. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

with clinically meaningful differences in outcomes – suggesting other sources of heterogeneity remain unexplored. One possibility, raised in recent literature, is that SMT trials using an A versus A + B design tend to report more short-term benefits, particularly in lower-quality studies (Grenier and Thiel, 2025) Still, the pervasive concerns observed, especially in more recent trials, limits the interpretability, reproducibility, and ultimately, credibility of SMT evidence.

4.1. Comparisons with the literature

Our findings reinforce prior evidence that reporting in SMT trials is inconsistent and incomplete (Sénéquier et al., 2025; Aspinall et al., 2024). Despite the availability of guidelines like the CONSORT (Moher et al., 2010; Begg et al., 1996) and CIRCLe-SMT (Groeneweg et al., 2017), adherence remains low. This ongoing lack of methodological detail impairs internal validity and limits the credibility of evidence used in systematic reviews (Nim et al., 2025; Rubinstein et al., 2019; Minnucci et al., 2023). Language and cultural barriers may also contribute to variability in how reporting guidelines are adopted outside English-speaking regions.

SMT trials appear less affected by publication integrity concerns, including retractions and suspected predatory practices, in comparison to trials on exercise for chronic spinal pain (Hayden et al., 2021, 2025). None of the included trials were retracted, most studies were published in journals that were members of a “verification-site”, and few studies showed evidence of very rapid acceptance. Sénéquier et al. reported that 23 % of osteopathy trials were accepted within two months, which they interpreted as a red flag for poor publication practices (Sénéquier et al., 2025). However, we used a different threshold, knowing that peer-review timelines vary widely depending on journal type and field. Fast acceptance (e.g., within two months) can be expected in many high-impact or specialty journals, making direct comparisons difficult (Andersen et al., 2021).

Comparing with Hayden et al. (2025) (Hayden et al., 2025), both studies identified four classes, from *Dated* to more recent trials with mixed reporting. Our *Non-contributing* class aligns closely with their *Research Waste* class, reflecting the continued publication of low-value trials in both SMT and exercise fields. While these trials display poor

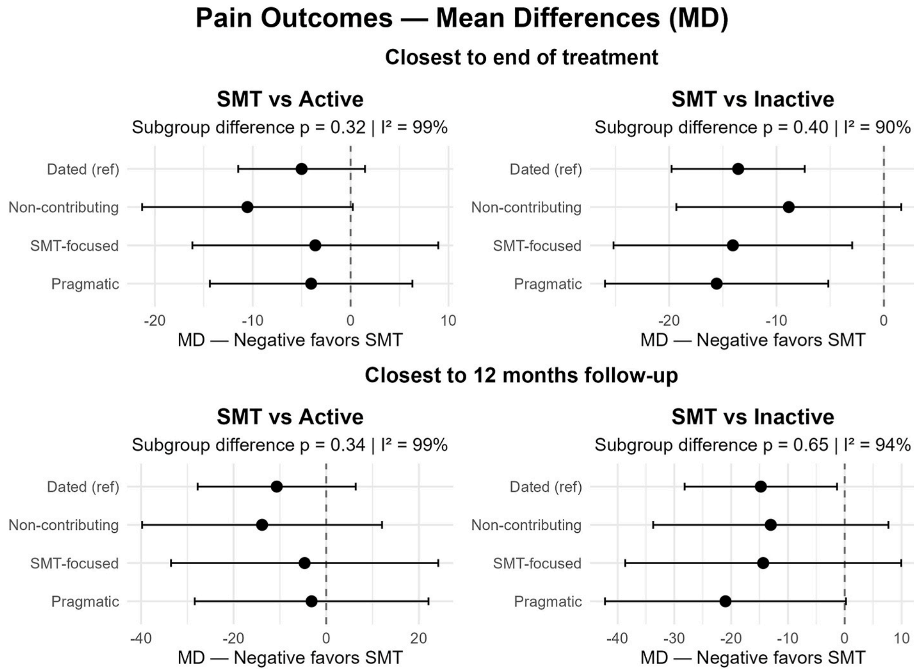

Fig. 4. Meta-analysis of pain outcomes by latent class.

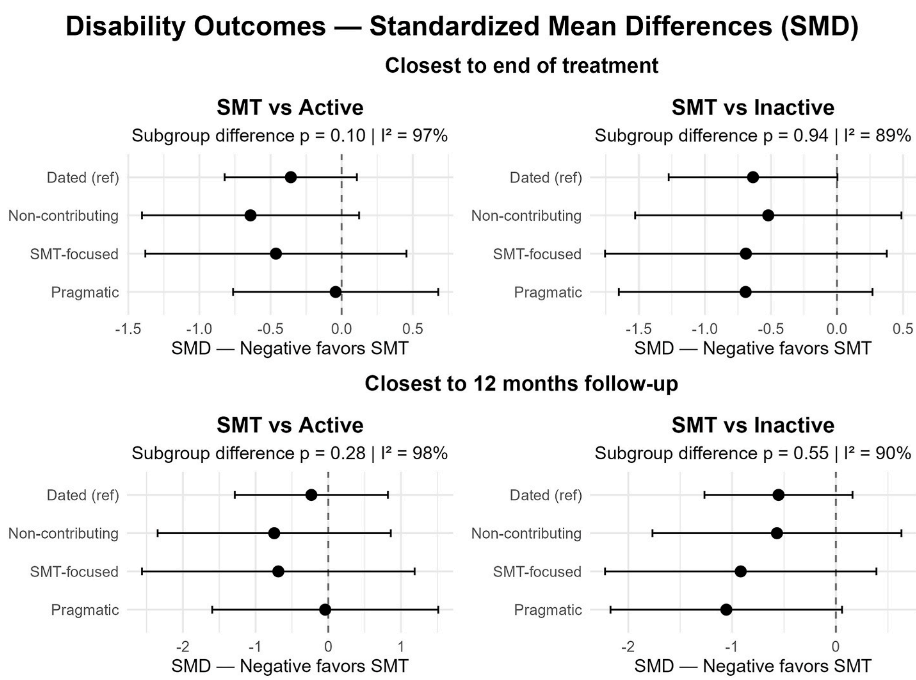

Fig. 5. Meta-analysis of disability outcomes by latent class.

reporting in many reporting domains, categorizing them solely as research waste may overlook their educational function (e.g., PhD or master theses), reflection on resource constraints in lower-income settings, requirement for larger grants, or perhaps have a more mechanistic focus. A unique contribution of our work was identifying the *SMT-focused* class, characterized by detailed fidelity reporting – an aspect not included in Hayden's framework. Finally, whereas Hayden observed a growing proportion of well-conducted trials, our findings suggest that "high-quality" SMT trials remain uncommon. Even within our *Pragmatic* class, reporting was less complete than in Hayden's *Well-conducted* group, underscoring a continued lag in reporting standards specifically within SMT literature.

4.2. Interpretation and implications

Across both pain and disability outcomes, the class-specific effect estimates showed some observable, but inconsistent, patterns that did not reach statistical significance. For comparisons with active interventions, the studies in the *Pragmatic* class tended to yield effect sizes closest to zero, suggesting that larger and more comprehensively reported trials produce more conservative estimates. In contrast, the *Non-contributing* class often produced the largest effect estimates in active controls, although the direction reversed in inactive controls, where this class showed among the smallest effects. The *Dated* class, despite poorer reporting across most domains, did not consistently show inflated effects and were generally more aligned with the *Pragmatic* class. These mixed patterns likely reflect the very high between-study heterogeneity and indicate that reporting and publication characteristics alone do not account for the variability in SMT trial outcomes.

Although class-specific estimates for pain and disability did not differ at a statistically significant level, improving transparency in SMT trial reporting and publication practices will help strengthen confidence, replicability, and interpretability of findings. This is particularly urgent as concerns grow over predatory publishing and paper mills (Candal-Pedreira et al., 2022; Soriano and Ruano-Ravina, 2025). While pooled effect estimates may appear stable, quality concerns make it difficult to judge the real-world relevance of findings. In line with the ENTRUST-PE framework, trust in pain evidence depends not only on statistical results but also on transparency and integrity in how trials are planned, conducted, and reported (O'Connell et al., 2025). Moving forward, editors and peer reviewers should more rigorously enforce established reporting standards, including CONSORT (with its Harms extension), TIDieR, and the CIRCLe SMT checklist (Aspinall et al., 2024).

4.3. Methodological considerations

A key strength of this study is the large sample derived from an up-to-date systematic review. Importantly, our results replicate the latent structure identified by Hayden et al. (2025) (Hayden et al., 2025), supporting the broader generalizability of these classes within clinical non-pharmacological spinal pain research.

However, several limitations must be considered. Our conclusions are based on what is reported in the papers. Hence, important methodological elements may have been well implemented in the trials but not reported, leading to potential misclassification. Although our indicators were selected based on theoretical and empirical rationales, some relevant features may be missing. For example, some variables used by Hayden et al. (2025) were not included in our model, and vice versa. Additionally, class labels reflect general patterns and do not capture every trial's unique profile. We did not weight class membership by posterior probability, but sensitivity analyses excluding studies with lower classification certainty (<0.8) yielded similar results.

Finally, the latent classes explained only a small fraction of the outcome variability. Thus, while reporting and publication integrity are important, other trial-level and clinical factors likely contribute more substantially to outcome heterogeneity.

5. Conclusion

In this meta-epidemiological study, we employed a data-driven approach to identify four classes of SMT RCTs based on trial reporting characteristics and publication practices. Despite incorporating more than 40 characteristics, class membership had minimal impact on estimates of treatment effects for pain and disability outcomes.

Our findings underscore the need for more rigorous adherence to established reporting guidelines to enhance the transparency, interpretability, and replicability of SMT research. We call for future research to focus explicitly on improving these integrity domains and to investigate their potential role in explaining outcome variability.

CRediT authorship contribution statement

Casper Nim: Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization. **Michelle Frederiksen:** Writing – review & editing, Methodology, Investigation. **Sasha Aspinall:** Writing – review & editing, Investigation. **Aron Downie:** Writing – review & editing, Investigation. **Martha Funabashi:** Writing – review & editing, Investigation. **Steen Harsted:** Writing – review & editing, Investigation. **Hazel Jenkins:** Writing – review & editing, Investigation. **David McNaughton:** Writing – review & editing, Investigation. **Luana Nyirö:** Writing – review & editing. **Eric J. Roseen:** Writing – review & editing, Investigation. **James J. Young:** Writing – review & editing, Investigation. **Liz Dennett:** Writing – review & editing, Methodology. **Stephen M. Perle:** Writing – review & editing. **Chad Cook:** Writing – review & editing, Methodology. **Carsten Juhl:** Writing – review & editing, Supervision, Methodology. **Jan Hartvigsen:** Writing – review & editing, Supervision, Methodology.

Registration

Registered in Open Science Framework, before data analysis (https://osf.io/xvphq/).

Funding

No funding was obtained.

Declaration of competing interest

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

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.msksp.2025.103480.

Data availability

Data used for the analysis are available upon reasonable request. Please contact corresponding author Casper Nim (casper.nim@rsyd.dk).

References

Alvarez, G., Solà, I., Sitjà-Rabert, M., Fort-Vanmeerhaeghe, A., Gich, I., Fernández, C., et al., 2020. A methodological review revealed that reporting of trials in manual therapy has not improved over time. J. Clin. Epidemiol. 121, 32–44.
Andersen, M.Z., Fonnes, S., Rosenberg, J., 2021. Time from submission to publication varied widely for biomedical journals: a systematic review. Curr. Med. Res. Opin. 37 (6), 985–993.
Aspinall, S.L., Nim, C., Hartvigsen, J., Cook, C.E., Skillgate, E., Vogel, S., et al., 2024. Waste not, want not: call to action for spinal manipulative therapy researchers. Chiropr. Man. Ther. 32 (1), 16.
Begg, C.B., Cho, M.K., Eastwood, S., Horton, R., Horton, R., Horton, R., et al., 1996. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. JAMA.
Bejarano, G., Csiernik, B., Zadro, J.R., Ferreira, G.E., 2023. Compared to what? An analysis of comparators in trials informing the National Institute of Clinical Excellence (NICE) low back pain guideline. Braz. J. Phys. Ther. 27 (6), 100563.
Bronfort, G., Goldsmith, C.H., Nelson, C.F., Boline, P.D., Anderson, A.V., 1996. Trunk exercise combined with spinal manipulative or NSAID therapy for chronic low back pain: a randomized, observer-blinded clinical trial. J. Manip. Physiol. Ther. 19 (9), 570–582.
Bronfort, G., Evans, R., Nelson, B., Aker, P.D., Goldsmith, C.H., Vernon, H., 2001. A randomized clinical trial of exercise and spinal manipulation for patients with chronic neck pain. Spine 26 (7), 788–797 discussion 798-799.
Bronfort, G., Evans, R., Anderson, A.V., Svendsen, K.H., Bracha, Y., Grimm, R.H., 2012. Spinal manipulation, medication, or home exercise with advice for acute and subacute neck pain: a randomized trial. Ann. Intern. Med. 156 (1 Pt 1), 1–10.
Candal-Pedreira, C., Ross, J.S., Ruano-Ravina, A., Egilman, D.S., Fernández, E., Pérez-Ríos, M., 2022. Retracted papers originating from paper mills: cross sectional study. BMJ 379, e071517.
Corp, N., Mansell, G., Stynes, S., Wynne-Jones, G., Morsø, L., Hill, J.C., et al., 2021. Evidence-based treatment recommendations for neck and low back pain across Europe: a systematic review of guidelines. Eur. J. Pain 25 (2), 275–295.
Eybye, M.N., Madsen, S.D., Schultz, A.N.Ø., Nim, C.G., 2022. Database coverage and their use in systematic reviews regarding spinal manipulative therapy: an exploratory study. Chiropr. Man. Ther. 30 (1), 57.
Ferreira, M.L., Luca, K de, Haile, L.M., Steinmetz, J.D., Culbreth, G.T., Cross, M., et al., 2023. Global, regional, and national burden of low back pain, 1990–2020, its attributable risk factors, and projections to 2050: a systematic analysis of the Global Burden of Disease Study 2021. Lancet Rheumatol 5 (6), e316–e329.

Foster, N.E., Anema, J.R., Cherkin, D., Chou, R., Cohen, S.P., Gross, D.P., et al., 2018. Prevention and treatment of low back pain: evidence, challenges, and promising directions. Lancet 391 (10137), 2368–2383.

Grenier, J., Thiel, A., 2025. Evaluating manual therapy in musculoskeletal pain: why certain trial designs may overestimate effectiveness—A scoping review. Eur J. Pain Lond Engl 29 (10), e70150.

Groeneweg, R., Rubinstein, S.M., Oostendorp, R.A.B., Ostelo, R.W.J.G., van Tulder, M. W., 2017. Guideline for reporting interventions on spinal manipulative therapy: consensus on interventions reporting criteria list for spinal manipulative therapy (CIRCLe SMT). J. Manip. Physiol. Ther. 40 (2), 61–70.

Hayden, J.A., Ellis, J., Ogilvie, R., Boulos, L., Stanojevic, S., 2021. Meta-epidemiological study of publication integrity, and quality of conduct and reporting of randomized trials included in a systematic review of low back pain. J. Clin. Epidemiol. 134, 65–78.

Hayden, J.A., Ogilvie, R., Kashif, S., Wilkinson, J., 2025. Identification of latent classes of randomized controlled trials based on integrity and reporting characteristics. J. Clin. Epidemiol. 179, 111644.

Higgins, J., Savović, J., Page, M., Elbers, R., Sterne, J., 2022. Cochrane Handbook for Systematic Reviews of Interventions (updated February 2022) [Internet]. Cochrane [cited 2023 Feb 21]. Available from:, version 6.3. www.training.cochrane.org/handbook.

Hurwitz, E.L., Morgenstern, H., Harber, P., Kominski, G.F., Belin, T.R., Yu, F., et al., 2002. A randomized trial of medical care with and without physical therapy and chiropractic care with and without physical modalities for patients with low back pain: 6-month follow-up outcomes from the UCLA low back pain study. Spine 27 (20), 2193–2204.

Linzer, D.A., Lewis, J.B., 2011. poLCA: an R package for polytomous variable latent class analysis. J. Stat. Software 42, 1–29.

McGee, D., Lorencatto, F., Matvienko-Sikar, K., Toomey, E., 2018. Surveying knowledge, practice and attitudes towards intervention fidelity within trials of complex healthcare interventions. Trials 19 (1), 504.

Minnucci, S., Innocenti, T., Salvioli, S., Giagio, S., Yousif, M.S., Riganelli, F., et al., 2023. Benefits and harms of spinal manipulative therapy for treating recent and persistent nonspecific neck pain: a systematic review with meta-analysis. J. Orthop. Sports Phys. Ther. 53 (9), 510–528.

Moher, D., Moher, D., Hopewell, S., Schulz, K.F., Montori, V.M., Gøtzsche, P.C., et al., 2010. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ.

Muñoz Laguna, J., Kurmann, A., Hofstetter, L., Nyantakyi, E., Braun, J., Clack, L., et al., 2025. "Which treatment do you believe you received?" A randomised blinding feasibility trial of spinal manual therapy. Chiropr. Man. Ther. 33 (1), 4.

Nim, C.G., Aspinall, S.L., Cook, C.E., Corrêa, L.A., Donaldson, M., Downie, A.S., et al., 2023. The effectiveness of spinal manipulative therapy procedures for spine pain: protocol for a systematic review and network meta-analysis. Chiropr. Man. Ther. 31 (1), 14.

Nim, C., Aspinall, S.L., Cook, C.E., Corrêa, L.A., Donaldson, M., Downie, A.S., et al., 2025. The effectiveness of spinal manipulative therapy in treating spinal pain does not depend on the application procedures: a systematic review and network meta-analysis. J. Orthop. Sports Phys. Ther. [cited 2025 Jan 20]; Available from: https://www.jospt.org/doi/10.2519/jospt.2025.12707.

O'Connell, N.E., Belton, J., Crombez, G., Eccleston, C., Fisher, E., Ferraro, M.C., et al., 2025. Enhancing the trustworthiness of pain research: a call to action. J. Pain [cited 2025 Aug 19];28. Available from: https://www.jpain.org/article/S1526-5900(24)00714-4/fulltext.

R Core Team, 2023. R: a Language and Environment for Statistical Computing [Internet]. R Foundation for Statistical Computing, Vienna, Austria. Available from. https://www.R-project.org/.

Rubinstein, S.M., de Zoete, A., van Middelkoop, M., Assendelft, W.J.J., de Boer, M.R., van Tulder, M.W., 2019. Benefits and harms of spinal manipulative therapy for the treatment of chronic low back pain: systematic review and meta-analysis of randomised controlled trials. BMJ [cited 2019 Jun 26];364. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6396088/.

Ruzich, J.J., Klopper, M., Dohrmann, C.M., Rhon, D.I., Young, J.L., 2024. How reproducible are manual therapy interventions in trials for low back pain? A scoping review. J. Orthop. Sports Phys. Ther. 54 (4), 248–257.

Sénéquier, A., Draper-Rodi, J., Alvarez Bustins, G., Braithwaite, F.A., Brown, J., Corcoran, D., et al., 2025. Investigating the trustworthiness of randomized controlled trials in osteopathic research: a systematic review with meta-analysis. J. Clin. Epidemiol. 183, 111788.

Snow, M., Trexler, S., Gates, C., Pudoka, K., Wilhelm, M., 2019. The fidelity of comparison intervention in manual therapy trial for patients with low back pain: a systematic review. Phys. Ther. Rev. 24 (5), 248–258.

Soriano, J.B., Ruano-Ravina, A., 2025. The rising threat of predatory journals and paper mills in respiratory medicine and research. Lancet Respir. Med. 13 (6), e30–e31.

UK BEAM Trial Team, 2004. United Kingdom back pain exercise and manipulation (UK BEAM) randomised trial: effectiveness of physical treatments for back pain in primary care. BMJ 329 (7479), 1377.

Viechtbauer, W., 2010. Conducting meta-analyses in R with the metafor package. J. Stat. Software 36 (3), 1–48.

Wewege, M.A., Jones, M.D., Williams, S.A., Kamper, S.J., McAuley, J.H., 2022. Rescaling pain intensity measures for meta-analyses of analgesic medicines for low back pain appears justified: an empirical examination from randomised trials. BMC Med. Res. Methodol. 22, 285.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., et al., 2019. Welcome to the tidyverse. J. Open Source Softw. 4 (43), 1686.

World Bank Country and Lending Groups. World Bank data help desk [Internet]. [cited 2025 Jun 2]. Available from: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519.

World Health Organization, 2023. WHO Guideline for Non-surgical Management of Chronic Primary Low Back Pain in Adults in Primary and Community Care Settings.

Wu, A.M., Cross, M., Elliott, J.M., Culbreth, G.T., Haile, L.M., Steinmetz, J.D., et al., 2024. Global, regional, and national burden of neck pain, 1990–2020, and projections to 2050: a systematic analysis of the Global Burden of Disease Study 2021. Lancet Rheumatol 6 (3), e142–e155.