International Journal of Osteopathic Medicine xxx (xxxx) xxx

Contents lists available at ScienceDirect

International Journal of Osteopathic Medicine

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

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

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

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

ARTICLE INFO

Keywords
CONSORT
Manual therapy
Reporting
P values
Randomized controlled trials

ABSTRACT

Background: The number of randomized clinical trials (RCTs) for manual therapy (MT) has increased exponentially in recent years but the quality of reporting is heterogeneous.
Objective: To assess the quality of the reporting of results in RCTs manual therapy, both in the text and in the graphs.
Study design: Methodological review.
Methods: We reviewed a random sample of 120 RCTs in MT published between 2000 and 2020 in indexed journals. We identified the primary outcome for each trial, and evaluated the completeness and correctness of reporting of results in the text and in the graphs.
Results: Forty per cent of the RCTs explicitly identified the primary outcome and 47.5% reported a sample size calculation. In 46.7% of the trials, the reporting of between groups comparisons was complete (including effect size and precision). Only 29.2% used the confidence interval as a measure of precision. Fifty-eight per cent of the trials reported significant differences in the results, and 30.8% reported a value of clinical relevance for at least one variable of the study. Forty-seven per cent reported the primary outcome graphically but only 19.6% of the graphs were self explanatory and 66.1% had problems of visual clarity.
Conclusions: Our findings suggest that the reporting of the results in MT trials is generally incomplete and graphics are often poor. These shortcomings could affect the interpretation of the results and their application in clinical practice. Improvements are needed in the reporting of results in order to advance clinical practice and research in manual therapy.

* Corresponding author. Department of Physical Therapy, Faculty of Medicine, University of Chile, Avenida Independencia 1027, Independencia, Santiago, Chile.
** Corresponding author. Iberoamerican Cochrane Center C, Sant Antoni Maria Claret 167 Pavelló 18, Planta 0 08025 Barcelona, Catalunya, Spain.
E-mail addresses: r_nunez@uchile.cl (R. Núñez-Cortés), gerardalv@gmail.com (G. Alvarez), javier.perez.b@gmail.com (J. Pérez-Bracchiglione), rosacabanas@uic.es (R. Cabanas-Valdés), jcalvosanz@gmail.com (J. Calvo-Sanz), XBonfill@santpau.cat (X. Bonfill), gurrutia@santpau.cat (G. Urrutia).
[1] RNC and GA contributed equally.
[2] Gerard Alvarez is a PhD student on Biomedical Research Methodology and Public Health in the Medical Department of the Universitat Autonoma de Barcelona, Barcelona, Spain.

https://doi.org/10.1016/j.ijosm.2021.06.002
Received 19 January 2021; Received in revised form 23 June 2021; Accepted 25 June 2021
Available online 12 July 2021
1746-0689/

Implications for practice

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

Introduction

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

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

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

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

Methods

Search strategy

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

Assessment of reporting

The reporting in the text

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

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

Graphic report

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

Statistical analysis

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

Results

The 120 trials included a total of 12954 participants (median = 60, range 6–1340) and they assessed the impact of six modalities of manual therapies: soft tissue techniques (29%); spinal manipulation techniques (23%); joint mobilization (21%); chiropractic treatments (12%); acupressure-reflexology (9%) and osteopathic manipulative treatment (6%). The most frequent type of trials included in the study were those related to spinal problems such as low back pain (21%) and neck pain (16%), followed by trials conducted in healthy or asymptomatic subjects (12.5%). Of the 120 articles analysed, 81% corresponded to single-centre studies and 19% to multi-centre studies. Treatment consisted of more than one session in 70% of the studies. Considering all categories, the agreement between pairs of reviewers in the pilot stage ranged from 0.737 to 0.853, which is interpreted as substantial to almost perfect agreement.

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

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

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

Table 2
Reporting specific information regarding results.

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

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

Analysis of the reporting of results in the text and tables

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

Analysis of the graphic report of the results

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

Discussion

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

Table 3
Comparison of the reporting of results by years.

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

Data are expressed in percentages (%).
[a] including alpha/beta values and standard deviation or minimum detectable change).
[b] including numbers, reasons and imputation criteria for missing data.
[c] including effect size and precision measure.

Table 4
Evaluation of the graphic report (n = 56).

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

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

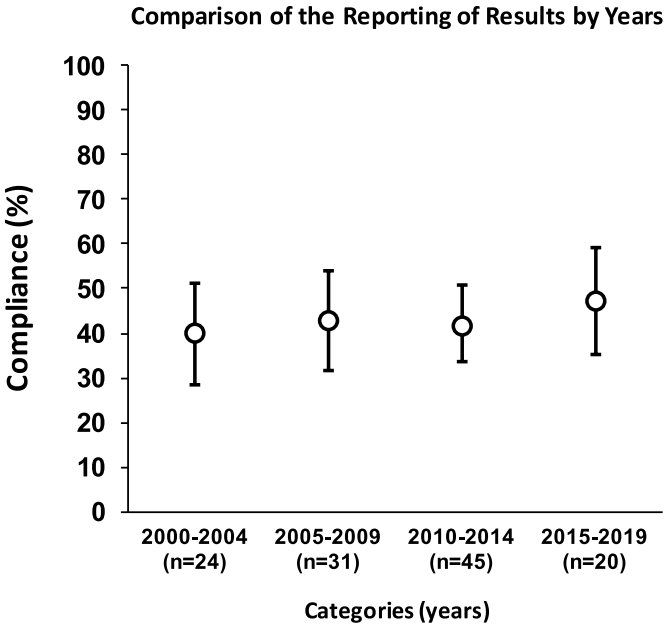

Fig. 1. Compliance (%) was defined by seven categories: a) Primary outcome is explicitly declared; b) Complete information on sample size calculation; c) Complete loss report; d) Intra-group effect measure (experimental group); e) Intra-group effect measure (control group); f) Complete comparative measure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

literature.

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

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

Primary and secondary outcomes

In our sample, fewer than 50% of the articles explicitly stated the primary outcome of the study. CONSORT guidelines have long recommended that the primary outcome should be specifically reported and differentiated from secondary outcomes [22]. The primary outcome measure is set at the moment the study is designed and the protocol is drafted. Outcomes are based directly on the primary aim of the study [45] and sample size is calculated from the primary outcome. Clearly stating the primary outcome prevents investigators from cherry-picking significant results and presenting these as the main findings of the study. To avoid this reporting bias, it is strongly recommended to prospectively register the study protocol on a public trial registry.

Sample size, statistical significance and clinical relevance

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

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

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

Effect estimates and precision

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

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

Loss to follow-up and missing data

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

Graphical report

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

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

Strengths and limitations

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

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

Conclusions

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

Author contributions

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

the manuscript. GU is the study guarantor.

Ethical approval

Not applicable.

Funding

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

Declaration of competing interest

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

Appendix 1. Data extraction form

Is the study hypothesis specified?
Yes, explicitly
No, but it is implicit
Not specified (cannot be determined)
If the hypothesis can be determined, it is:
Superiority
Non-inferiority
Is a primary variable described?
Yes (explicitly specified)
No (but it can be identified from the sample size calculation)
No
Variable type (n = 100)
Continuous
Dichotomous
Ordinal category
Follow-up (n = 100)
Short term (≤ 3 months)
Long term (> 3 months)
Statistical analysis (n = 100)
ANOVA
Mixed repeated measures model
Mean difference
T test
Chi-Square
Risk Ratio, Odds Ratio or Hazard Ratio
ANCOVA
Mann–Whitney U test
Regression models
Other
Was the result for the main variable presented graphically?
Yes
No
Was a sample size calculation performed?
Yes
No
What information is given to evaluate how this calculation was made?
Alpha value
Beta value
Minimal change to be achieved with the intervention
Standard deviation in continuous variables
Expected rate or mean in the control group
Delta margin (for non-inferiority studies)
Is there sufficient information to determine whether there was a loss of subjects during the study?
Yes
No
In the event that some subjects were lost to follow-up, did they appear quantified (n°) in the article?
Yes
No
Not applicable (no losses)
If some subjects were lost to follow-up, were reasons described?
Yes
No
Not applicable (no losses)
If some subjects were lost to follow up, were the criteria for imputation of these missing data described?
Yes, analysis by protocol
Yes, other reasons
No
Not applicable (no losses)
Is the measure of the intra-group effect reported?
Measure of effect is reported
Only data from each evaluation is reported

(*continued on next page*)

(*continued*)

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

References

[1] French HP, Brennan A, White B, Cusack T. Manual therapy for osteoarthritis of the hip or knee - a systematic review. Man Ther 2011;16:109–17.
[2] Alvarez G, Solà I, Sitjà-Rabert M, Fort-Vanmeerhaeghe A, Gich I, Fernández C, et al. A methodological review revealed that reporting of trials in manual therapy has not improved over time. J Clin Epidemiol 2020;121:32–44. https://doi.org/10.1016/j.jclinepi.2020.01.006.
[3] Shiwa SR, Moseley AM, Maher CG, Costa LOP. Language of publication has a small influence on the quality of reports of controlled trials of physiotherapy interventions. J Clin Epidemiol 2013;66:78–84. https://doi.org/10.1016/j.jclinepi.2012.08.004.
[4] Zavala-González J, Pavez-Baeza F, Gutiérrez-Espinoza H, Olguín-Huerta C. The effectiveness of joint mobilization techniques for range of motion in adult patients with primary adhesive capsulitis of the shoulder: a systematic review and meta-analysis. Medwave 2018;18:e7265.
[5] Minkalis AL, Vining RD, Long CR, Hawk C, de Luca K. A systematic review of thrust manipulation combined with one conservative intervention for rotator cuff and related non-surgical shoulder conditions. J Can Chiropr Assoc 2018;62:5–17.
[6] Carnes D, Plunkett A, Ellwood J, Miles C. Manual therapy for unsettled, distressed and excessively crying infants: a systematic review and meta-analyses. BMJ Open 2018;8:e019040.
[7] de Luca KE, Fang SH, Ong J, Shin K-S, Woods S, Tuchin PJ. The effectiveness and safety of manual therapy on pain and disability in older persons with chronic low back pain: a systematic review. J Manip Physiol Ther 2017;40:527–34.
[8] Bizzarri P, Buzzatti L, Cattrysse E, Scafoglieri A. Thoracic manual therapy is not more effective than placebo thoracic manual therapy in patients with shoulder dysfunctions: a systematic review with meta-analysis. Musculoskeletal Sci Pract 2018;33:1–10.
[9] Laimi K, Mäkilä A, Bärlund E, Katajapuu N, Oksanen A, Seikkula V, et al. Effectiveness of myofascial release in treatment of chronic musculoskeletal pain: a systematic review. Clin Rehabil 2018;32:440–50.
[10] Wu J, Wang Y, Wang Z. The effectiveness of massage and touch on behavioural and psychological symptoms of dementia: a quantitative systematic review and meta-analysis. J Adv Nurs 2017;73:2283–95.
[11] Smith CA, Levett KM, Collins CT, Dahlen HG, Ee CC, Suganuma M. Massage, reflexology and other manual methods for pain management in labour. Cochrane Database Syst Rev 2018;3:CD009290.
[12] Furlan AD, Giraldo M, Baskwill A, Irvin E, Imamura M. Massage for low-back pain. Cochrane Database Syst Rev 2015. https://doi.org/10.1002/14651858.cd001929.pub3.
[13] Westad K, Tjoestolvsen F, Hebron C. The effectiveness of Mulligan's mobilisation with movement (MWM) on peripheral joints in musculoskeletal (MSK) conditions: a systematic review. Musculoskeletal Sci Pract 2018. https://doi.org/10.1016/j.msksp.2018.12.001.
[14] Lee N-W, Kim G-H, Heo I, Kim K-W, Ha I-H, Lee J-H, et al. Chuna (or tuina) manual therapy for musculoskeletal disorders: a systematic review and meta-analysis of randomized controlled trials. Evid Based Compl Alternat Med 2017:8218139. 2017.
[15] Xu Q, Chen B, Wang Y, Wang X, Han D, Ding D, et al. The effectiveness of manual therapy for relieving pain, stiffness, and dysfunction in knee osteoarthritis: a systematic review and meta-analysis. Pain Physician 2017;20:229–43.
[16] Mischke JJ, Jayaseelan DJ, Sault JD, Emerson Kavchak AJ. The symptomatic and functional effects of manual physical therapy on plantar heel pain: a systematic review. J Man Manip Ther 2017;25:3–10.
[17] Krul M, van der Wouden JC, Kruithof EJ, van Suijlekom-Smit LW, Koes BW. Manipulative interventions for reducing pulled elbow in young children. Cochrane Database Syst Rev 2017;7:CD007759.
[18] Gross A, Langevin P, Burnie SJ, Bédard-Brochu M-S, Empey B, Dugas E, et al. Manipulation and mobilisation for neck pain contrasted against an inactive control or another active treatment. Cochrane Database Syst Rev 2015. https://doi.org/10.1002/14651858.cd004249.pub4.
[19] Silva TM, Costa LCM, Costa LOP. Evidence-Based Practice: a survey regarding behavior, knowledge, skills, resources, opinions and perceived barriers of Brazilian physical therapists from São Paulo state. Braz J Phys Ther 2015;19:294–303.
[20] da Silva TM, Costa L da CM, Garcia AN, Costa LOP. What do physical therapists think about evidence-based practice? A systematic review. Man Ther 2015;20:388–401.
[21] Straulino S. Reconstruction of Galileo Galilei's experiment: the inclined plane. Phys Educ 2008;43:316–21. https://doi.org/10.1088/0031-9120/43/3/012.
[22] Altman DG. Better reporting of randomised controlled trials: the CONSORT statement. BMJ 1996;313:570–1.
[23] Caulley L, Catalá-López F, Whelan J, Khoury M, Ferraro J, Cheng W, et al. Reporting guidelines of health research studies are frequently used inappropriately. J Clin Epidemiol 2020;122:87–94.
[24] Moseley AM, Elkins MR, Janer-Duncan L, Hush JM. The quality of reports of randomized controlled trials varies between subdisciplines of physiotherapy. Physiother Can 2014;66:36–43.
[25] Moseley AM, Herbert RD, Maher CG, Sherrington C, Elkins MR. Reported quality of randomized controlled trials of physiotherapy interventions has improved over time. J Clin Epidemiol 2011;64:594–601.
[26] Riley SP, Swanson B, Brismée J-M, Sawyer SF. A systematic review of orthopaedic manual therapy randomized clinical trials quality. J Man Manip Ther 2016;24:241–52.
[27] Puhan MA, ter Riet G, Eichler K, Steurer J, Bachmann LM. More medical journals should inform their contributors about three key principles of graph construction. J Clin Epidemiol 2006;59:1017–22.
[28] Pocock SJ, Travison TG, Wruck LM. Figures in clinical trial reports: current practice & scope for improvement. Trials 2007;8:36.
[29] Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Intern Med 2001;134:663–94.
[30] Pillastrini P, Vanti C, Curti S, Mattioli S, Ferrari S, Violante FS, et al. Using PubMed search strings for efficient retrieval of manual therapy research literature. J Manip Physiol Ther 2015;38:159–66.
[31] Butcher NJ, Monsour A, Mew EJ, Szatmari P, Pierro A, Kelly LE, et al. Improving outcome reporting in clinical trial reports and protocols: study protocol for the Instrument for reporting Planned Endpoints in Clinical Trials (InsPECT). Trials 2019;20:161.

[32] Butcher NJ, Mew EJ, Saeed L, Monsour A, Chee-A-Tow A, Chan A-W, et al. Guidance for reporting outcomes in clinical trials: scoping review protocol. BMJ Open 2019;9:e023001.
[33] Chen JC, Cooper RJ, McMullen ME, Schriger DL. Graph quality in top medical journals. Ann Emerg Med 2017;69:453. 61.e5.
[34] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.
[35] Bhaloo Z, Adams D, Liu Y, Hansraj N, Hartling L, Terwee CB, et al. Primary Outcomes Reporting in Trials (PORTal): a systematic review of inadequate reporting in pediatric randomized controlled trials. J Clin Epidemiol 2017;81:33–41.
[36] Sinha IP, Altman DG, Beresford MW, Boers M, Clarke M, Craig J, et al. Standard 5: selection, measurement, and reporting of outcomes in clinical trials in children. Pediatrics 2012;129(Suppl 3):S146–52.
[37] Mayo-Wilson E, Fusco N, Li T, Hong H, Canner JK, Dickersin K, et al. Multiple outcomes and analyses in clinical trials create challenges for interpretation and research synthesis. J Clin Epidemiol 2017;86:39–50.
[38] Mantziari S, Demartines N. Poor outcome reporting in medical research; building practice on spoilt grounds. Ann Transl Med 2017;5:S15.
[39] Dwan K, Gamble C, Williamson PR, Kirkham JJ. Reporting Bias Group. Systematic review of the empirical evidence of study publication bias and outcome reporting bias - an updated review. PloS One 2013;8:e66844.
[40] Hall NJ, Kapadia MZ, Eaton S, Chan WWY, Nickel C, Pierro A, et al. Outcome reporting in randomised controlled trials and meta-analyses of appendicitis treatments in children: a systematic review. Trials 2015;16:275.
[41] Saldanha IJ, Dickersin K, Wang X, Li T. Outcomes in Cochrane systematic reviews addressing four common eye conditions: an evaluation of completeness and comparability. PloS One 2014;9:e109400.
[42] Chan A-W, Altman DG. Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors. BMJ 2005;330:753.
[43] Froud R, Underwood M, Eldridge S. Improving the reporting and interpretation of clinical trial outcomes. Br J Gen Pract 2012;62:e729–31. https://doi.org/10.3399/bjgp12x657008.
[44] Heneghan C, Goldacre B, Mahtani KR. Why clinical trial outcomes fail to translate into benefits for patients. Trials 2017;18:122.
[45] Vetter TR, Mascha EJ. Defining the primary outcomes and justifying secondary outcomes of a study: usually, the fewer, the better. Anesth Analg 2017;125:678–81.
[46] Abdul Latif L, Daud Amadera JE, Pimentel D, Pimentel T, Fregni F. Sample size calculation in physical medicine and rehabilitation: a systematic review of reporting, characteristics, and results in randomized controlled trials. Arch Phys Med Rehabil 2011;92:306–15.
[47] Castellini G, Gianola S, Bonovas S, Moja L. Improving power and sample size calculation in rehabilitation trial reports: a methodological assessment. Arch Phys Med Rehabil 2016;97:1195–201.
[48] Koes BW. How to evaluate manual therapy: value and pitfalls of randomized clinical trials. Man Ther 2004;9:183–4.
[49] Rubinstein SM, van Eekelen R, Oosterhuis T, de Boer MR, Ostelo RWJG, van Tulder MW. The risk of bias and sample size of trials of spinal manipulative therapy for low back and neck pain: analysis and recommendations. J Manip Physiol Ther 2014;37:523–41.
[50] Page P. Beyond statistical significance: clinical interpretation of rehabilitation research literature. Int J Sports Phys Ther 2014;9:726–36.
[51] Armijo-Olivo S. The importance of determining the clinical significance of research results in physical therapy clinical research. Braz J Phys Ther 2018;22:175–6.
[52] Armijo-Olivo S, Warren S, Fuentes J, Magee DJ. Clinical relevance vs. statistical significance: using neck outcomes in patients with temporomandibular disorders as an example. Man Ther 2011;16:563–72.
[53] Rossettini G, Carlino E, Testa M. Clinical relevance of contextual factors as triggers of placebo and nocebo effects in musculoskeletal pain. BMC Muscoskel Disord 2018;19. https://doi.org/10.1186/s12891-018-1943-8.
[54] Bialosky JE, Beneciuk JM, Bishop MD, Coronado RA, Penza CW, Simon CB, et al. Unraveling the mechanisms of manual therapy: modeling an approach. J Orthop Sports Phys Ther 2018;48:8–18.
[55] Bialosky JE, Bishop MD, Penza CW. Placebo mechanisms of manual therapy: a sheep in wolf's clothing? J Orthop Sports Phys Ther 2017;47:301–4.
[56] Kamper SJ. Interpreting outcomes 1—change and difference: linking evidence to practice. J Orthop Sports Phys Ther 2019;49:357–8.
[57] Goodman S. A dirty dozen: twelve P-value misconceptions. Semin Hematol 2008;45:135–40. https://doi.org/10.1053/j.seminhematol.2008.04.003.
[58] Oliveri RS, Gluud C, Wille-Jørgensen PA. Hospital doctors' self-rated skills in and use of evidence-based medicine - a questionnaire survey. J Eval Clin Pract 2004;10:219–26.
[59] Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, et al. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. Eur J Epidemiol 2016;31:337–50.
[60] Bassler D, Briel M, Montori VM, Lane M, Glasziou P, Zhou Q, et al. Stopping randomized trials early for benefit and estimation of treatment effects: systematic review and meta-regression analysis. J Am Med Assoc 2010;303:1180–7.
[61] Bleakley C, Reijgers J, Smoliga JM. Many high-quality randomized controlled trials in sports physical therapy are making false-positive claims of treatment effect: a systematic survey. J Orthop Sports Phys Ther 2020;50:104–9. https://doi.org/10.2519/jospt.2020.9264.
[62] Fethney J. Statistical and clinical significance, and how to use confidence intervals to help interpret both. Aust Crit Care 2010;23:93–7.
[63] Kamper SJ. Confidence intervals: linking evidence to practice. J Orthop Sports Phys Ther 2019;49:763–4.
[64] Armijo-Olivo S, Warren S, Magee D. Intention to treat analysis, compliance, drop-outs and how to deal with missing data in clinical research: a review. Phys Ther Rev 2009;14:36–49.
[65] von Allmen RS, Weiss S, Tevaearai HT, Kuemmerli C, Tinner C, Carrel TP, et al. Completeness of follow-up determines validity of study findings: results of a prospective repeated measures cohort study. PloS One 2015;10:e0140817.
[66] Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ 2010;340:c869. 10 1136/bmj.c869.
[67] Akl EA, Briel M, You JJ, Sun X, Johnston BC, Busse JW, et al. Potential impact on estimated treatment effects of information lost to follow-up in randomised controlled trials (LOST-IT): systematic review. BMJ 2012;344:e2809.
[68] Schriger DL, Sinha R, Schroter S, Liu PY, Altman DG. From submission to publication: a retrospective review of the tables and figures in a cohort of randomized controlled trials submitted to the British Medical Journal. Ann Emerg Med 2006;48. 750–6, 756.e1–21.
[69] Schriger DL, Cooper RJ. Achieving graphical excellence: suggestions and methods for creating high-quality visual displays of experimental data. Ann Emerg Med 2001;37:75–87.
[70] Pocock SJ, Travison TG, Wruck LM. How to interpret figures in reports of clinical trials. BMJ 2008;336:1166–9.
[71] Cleveland WS. The elements of graphing data. 1992.
[72] Blanco D, Altman D, Moher D, Boutron I, Kirkham JJ, Cobo E. Scoping review on interventions to improve adherence to reporting guidelines in health research. BMJ Open 2019;9:e026589.
[73] Hopewell S, Boutron I, Altman DG, Barbour G, Moher D, Montori V, et al. Impact of a web-based tool (WebCONSORT) to improve the reporting of randomised trials: results of a randomised controlled trial. BMC Med 2016;14:199.
[74] Barnes C, Boutron I, Giraudeau B, Porcher R, Altman DG, Ravaud P. Impact of an online writing aid tool for writing a randomized trial report: the COBWEB (Consort-based WEB tool) randomized controlled trial. BMC Med 2015;13:221.