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

Gerard Alvarez,[a,b] Rodrigo Núñez-Cortés,[c] Ivan Solà,[a,d] Mercè Sitjà-Rabert,[e,f] Azahara Fort-Vanmeerhaeghe,[e,g] Carles Fernández,[e,f] Xavier Bonfill,[a,d] Gerard Urrutia[a,d]

Affiliations:

a. Iberoamerican Cochrane Centre – Sant Pau Biomedical Research Institute, IIB Sant Pau, Barcelona, Spain
b. Foundation Centre for Osteopathic Medicine Collaboration. Spain National Centre, Barcelona, Spain
c. Department of Physical Therapy, Faculty of Medicine, University of Chile, Santiago, Chile
d. CIBER Epidemiología y Salud Pública (CIBERESP), Spain
e. Blanquerna School of Health Science (FCS), Ramon Llull University, Barcelona, Spain
f. Global Research on Wellbeing (GRoW) Research Group, Ramon Llull University, Barcelona, Spain
g. Blanquerna Faculty of Psychology, Education Sciences and Sport (FPCEE), Ramon Llull University, Barcelona, Spain

Contact author:	**1st Author's details:**
Gerard Alvarez: gerardalv@gmail.com +34 620 214 206 Iberoamerican Cochrane Center C. Sant Antoni Maria Claret 167 Pavelló 18 | Planta 0 08025 Barcelona | Catalunya | Spain	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.

Gerard Alvarez (gerardalv@gmail.com)
Iberoamerican Cochrane Centre
Sant Pau Biomedical Research Institute (IIB Sant Pau), Barcelona, Spain
Foundation Centre for Osteopathic Medicine Collaboration. Spain National Centre, Barcelona, Spain

Rodrigo Núñez-Cortés (rod.nunezcortes@gmail.com)
Department of Physical Therapy, Faculty of Medicine, University of Chile, Santiago, Chile.

Ivan Solà (ISola@santpau.cat)
Iberoamerican Cochrane Center
Sant Pau Biomedical Research Institute (IIB Sant Pau), Barcelona, Spain
CIBER Epidemiología y Salud Pública (CIBERESP), Spain

Mercè Sitjà-Rabert (mercesr@blanquerna.url.edu)
Blanquerna School of Health Science (FCS), Ramon Llull University, Barcelona, Spain
Global Research on Wellbeing (GRoW) Research Group, Ramon Llull University, Barcelona, Spain

Azahara Fort-Vanmeerhaeghe (azaharafv@blanquerna.url.edu)
Blanquerna School of Health Science (FCS), Ramon Llull University, Barcelona, Spain
Blanquerna Faculty of Psychology, Education Sciences and Sport (FPCEE), Ramon Llull University, Barcelona, Spain

Carles Fernández Jané (carlesfj@blanquerna.url.edu)
Blanquerna School of Health Science (FCS), Ramon Llull University, Barcelona, Spain
Global Research on Wellbeing (GRoW) Research Group, Ramon Llull University, Barcelona, Spain

Xavier Bonfill (XBonfill@santpau.cat)
Iberoamerican Cochrane Centre
Biomedical Research Institute Sant Pau (IIB Sant Pau), Barcelona, Spain
CIBER Epidemiología y Salud Pública (CIBERESP), Spain

Gerard Urrutia (GUrrutia@santpau.cat)
Iberoamerican Cochrane Centre
Sant Pau Biomedical Research Institute (IIB Sant Pau), Barcelona, Spain
CIBER Epidemiología y Salud Pública (CIBERESP), Spain

ABSTRACT

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

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

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

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

What is new?

Key Findings

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

What this adds to what was known?

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

What is the implication and what should change now?

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

Keywords: Self-acknowledged limitations, manual therapy, reporting, quality, transparency, guidelines

1. INTRODUCTION

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

This notwithstanding, insufficient acknowledgement of limitations in discussion sections has been previously found to be among the most important deficiencies in scientific articles[4]. Some studies have reported a lack of or inappropriate recognition of limitations in the medical literature. A study by Ioannidis, using automated key-word searching, found that only 17% of 400 papers published in leading medical journals used at least one word referring to limitations[5]. Ter Riet et al. analyzed the reporting of limitations and the use of "hedging patterns" (i.e. modal verbs like "may" or "could", adverbs like "apparently" or "possibly", and lexical verbs like "suggest [that]" or "believe [that]") in 300 biomedical publications published in 30 high and medium-ranked journals in 2007. They found that 27% of the articles did not report any limitations, and that there were major differences between journals in how uncertainty was expressed and limitations acknowledged. Despite improvement over time, the conclusion was that the reporting of limitations was probably incomplete[6]. Collectively, this literature shows that failing to acknowledge limitations is still a significant problem in the biomedical literature. Proposed strategies to help improve transparency and contribute to the evolving culture of research reproducibility include biomedical text mining to automatically recognize self-acknowledged limitations (SALs) in clinical research publications and "second independent" discussions (the participation of an additional author who is independent to the study and whose role is to identify its limitations)[7–9].

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

reflect space availability[5]. The increasing existence of open-access journals and online publishing should be seen as a solution regarding length restrictions.

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

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

2. METHODS

2.1 Eligibility criteria

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

2.1 Data extraction

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

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

2.2 Data analysis

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

3. RESULTS

3.1 Description of articles

The 120 trials analyzed assessed the impact of six different modalities of manual therapies *(soft tissue techniques (29%), spinal manipulation therapy (23%), joint mobilization (21%), chiropractic treatments (12%), acupressure/reflexology (10%) and osteopathic manipulative treatment (5%))* comparing them with up to 20 different types of control treatments/interventions. In some studies, more than one control intervention was used, and 'usual care' (21%), sham interventions (19%) and exercise (18%) were the most common. The trials included a mean of 111 participants (SD=187). The most frequent studies evaluated physiological responses on healthy or asymptomatic subjects and the most frequent health problems assessed were lower back pain and neck pain. Thirty-nine percent of the articles reported long-term follow-up. We reported a full description of the articles included elsewhere[15].
The articles studied were published in journals that ranged from 0.12 to 4.63 in the Scientific Journal Ranking (Scimago). At the year of publication, 74% of them were first quartile journals, 20% were second quartile and 6% were third quartile[25].

3.2 Number of limitations

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

3.3 Categories and specific limitations reported

Table 2 shows the frequency and categories of SALs reported in the articles. The most common limitation declared, in almost half of our sample, related to sample size (47.5%), followed by limitations related to study length and follow-up (33.3%) and inadequate controls (32.5%). Limitations regarding funding or

the selected centers and/or care providers were the least reported. For each category of limitations, several explanations were given by the authors, and in some cases, they reported different study weaknesses leading to the same type of specific limitation. **Table 2** shows the frequency of these explanations, which allow better understanding of the specific limitations within each category. For example, regarding sample size limitations, in 87.7% of the articles, authors declared a lack of statistical power to provide reliable results, while just 8.8% declared recruitment difficulties.

4. DISCUSSION

In the present study, we quantified and analyzed SALs in a sample of 120 manual therapy trials. While previous research assessing SALs has focused on high-impact journals regardless the study field[5,6], we extended our focus to trials published in any journal to capture the breadth of this kind of clinical research in manual therapy. Overall, our results indicate that at least two different limitations are consistently acknowledged in MT reports. Furthermore, the analysis of the arguments reported by authors provide some explanation to help understand the specific difficulties that MT investigators face and in turn establish strategies, recommendations or guidelines to improve these. The following discussion focuses on those SALs present in at least 20% of our sample.

4.1 Limitations related to sample size

The most reported SAL related to sample size (47.5%) and, within this category, the main reason given by the authors was that the studies lacked statistical power to detect effect differences between groups (87.7%). In general, RCTs in MT frequently involve small sample sizes[26]. Research in MT is often conducted in private centers and outpatient services which limits recruitment possibilities, leading to underpowered studies. For example, in our sample, although pilot studies were not excluded (n=15), only 31% of the RCTs had samples greater than 100 participants[15].

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

While adequate sample size is desirable, some authors have argued that underpowered trials might be acceptable if investigators use methodological rigor to eliminate bias, report properly to avoid misinterpretation, and always publish results to avoid publication bias[30]. According to Schulz et al., readers

should be more concerned about systematic errors rather than inadequate sample size. In fact, sample size deals more with the precision of the estimation of effect and not necessarily with the validity of the study[26]. The power of the trial is expressed in the confidence interval. Hence, if methodological issues are carefully considered, the power is no longer a major concern[30]. In turn, validity can dramatically influence the effect size. For example, in the field of low back pain, studies with higher risk of bias reported effect sizes that were, on average, 50% greater than the estimates reported in trials with a lower risk of bias[31]. In a sample of trials evaluating treatments for hip and knee osteoarthritis, it has been shown that inadequate random sequence generation and lack of allocation concealment and double-blinding yielded larger treatment effects[20].

4.2 Limitations related to inadequate control intervention

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

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

Template for Intervention Description and Replication (TIDieR) for placebo-controlled trials (TIDieR – Placebo) is currently registered in the Equator Network as a guideline under development[41], which is expected to be a useful tool to facilitate the design of better research in the field.

4.3 Limitations related to blinding

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

4.4 Limitations related to study length and follow-up

Study length and follow-up is another common limitation in MT trials (33.3% in our sample). The main study length limitations declared related to the lack of a long-term follow-up (87.5%). Collectively, the literature on specific mechanisms of MT concludes that the main effects are short-term[36,37]. However, in clinical practice, MT is commonly used to treat chronic conditions[45,52] through therapeutic packages that can last several weeks or months. Therefore, long-term follow-up should be encouraged in MT clinical trials assessing chronic conditions. Ideally, MT investigators should develop intervention models that match the natural history of common problems, considering the trajectory of the symptoms over time as an outcome rather than focusing on specified follow-up time points alone[53].

4.5 Limitations related to inadequate subject selection

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

4.6 Limitations related to the intervention

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

and reporting of some interventions. The reporting of non-pharmacological interventions is poor and incomplete[22], and that includes all fields of physical therapy interventions[60] and specifically MT[13,15]. However, beyond the reporting, there is a need to recognize the particularities of patient-centered touch-based interventions and strike a balance between internal and external validity. Some of the most recent reporting guidelines (e.g. the updated CONSORT extension for non-pharmacological interventions (CONOSRTnpt)[61] or the Template for Interventions Description and Replication (TIDieR))[22] already include items related to the tailoring of interventions. However, recent research shows that despite the existence of such guidelines, clinical reproducibility is compromised particularly for complex interventions and, as we proposed in a previous publication[15], there is a need for field-specific checklists[59]. Examples of this are the specific reporting guidelines on spinal manipulative therapy (SMT)[62] and therapeutic exercise[63].

4.7 Limitations related to outcome measures

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

4.8 Comment on reported limitations

In the previous sections, we have quantified and described those SALs identified in our sample. However, beyond the descriptive purposes of the present study, this issue deserves further comment in order to discuss SALs nature, their implications and impact. Although reported in a different proportion, we could

differentiate some SALs with a greater impact on the internal trial validity from others not directly linked to the study design or that reflected issues related to its conducting. Within the first group eight trials acknowledged potential selection bias due to inconsistencies in the randomization/concealment process, three trials acknowledged unequal follow-up between groups and four acknowledged suboptimal statistical analysis. In the second group, we identified SALs that were not truly methodological (e.g. limited funding in five trials) or reported problems derived from the trial progress (e.g. unbalanced groups in two trials) which should be amendable with a rigorous and accurate data analysis.

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

4.9 Study limitations

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

5. CONCLUSION

Recognition and discussion of all potentially important limitations of a study represent a crucial part of the scientific discourse. Our results indicate that at least two different limitations are consistently acknowledged in MT studies, and that those related to sample size, study length and follow-up and inadequate controls were the most commonly reported. Besides the items already included in reporting guidelines, specific extensions on how to report and acknowledge limitations could potentially enhance research quality and transparency.

Furthermore, researchers should be encouraged to report those limitations that could entail a threat to their studies validity. Systematic analysis of the reasons behind the SALs provides some insights on the main difficulties in conducting studies in this field and may prove helpful in the development of strategies to improve future research.

ACKNOWLEDGEMENTS

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

FUNDING SOURCES

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

REFERENCES

1. Price Editor JH, Murnan Assistant J. Research Limitations and the Necessity of Reporting Them. *Am J Heal Educ*. 2013;35(2):66-67.
2. Knottnerus JA, Tugwell P. Improving the clinical significance of medical research. *J Clin Epidemiol*. 2014;67(6):607-608.
3. Puhan MA, Akl EA, Bryant D, Xie F, Apolone G, ter Riet G. Discussing study limitations in reports of biomedical studies- the need for more transparency. *Health Qual Life Outcomes*. 2012;10(1):23.
4. Goodman SN, Berlin J, Fletcher SW, Fletcher RH. Manuscript quality before and after peer review and editing at Annals of Internal Medicine. *Ann Intern Med*. 1994;121(1):11-21.
5. Ioannidis JPA. Limitations are not properly acknowledged in the scientific literature. *J Clin Epidemiol*. 2007;60(4):324-329.
6. ter Riet G, Chesley P, Gross AG, et al. All That Glitters Isn't Gold: A Survey on Acknowledgment of Limitations in Biomedical Studies. *PLoS One*. 2013;8(11):e73623--6.
7. Kilicoglu H. Biomedical text mining for research rigor and integrity: tasks, challenges, directions. *Brief Bioinform*. 2017;533(9900):415-452.
8. Kilicoglu H, Rosemblat G, Malički M, ter Riet G. Automatic recognition of self-acknowledged limitations in clinical research literature. *J Am Med Informatics Assoc*. 2018;25(7):855-861.
9. Avidan MS, Ioannidis JPA, Mashour GA. Independent discussion sections for improving inferential reproducibility in published research. *Br J Anaesth*. 2019;122(4):413-420.
10. French HP, Brennan A, White B, Cusack T. Manual therapy for osteoarthritis of the hip or knee - A systematic review. *Man Ther*. 2011;16(2):109-117. doi:10.1016/j.math.2010.10.011
11. Smith AR. Manual therapy: The historical, current, and future role in the treatment of pain. *ScientificWorldJournal*. 2007;7:109-120. doi:10.1100/tsw.2007.14

12. Moseley AM, Elkins MR, Janer-Duncan L, Hush JM. The quality of reports of randomized controlled trials varies between subdisciplines of physiotherapy. *Physiother Canada*. 2014;66(1):36-43. doi:10.3138/ptc.2012-68
13. Riley SP, Swanson B, Brismée J-M, Sawyer SF. A systematic review of orthopaedic manual therapy randomized clinical trials quality. *J Man Manip Ther*. 2016;24(5):241-252. doi:10.1080/10669817.2015.1119372
14. Gonzalez GZ, Moseley AM, Maher CG, Nascimento DP, Costa L i ola da CMLOL da CM, Costa L i ola da CMLOL da CM. Methodologic Quality and Statistical Reporting of Physical Therapy Randomized Controlled Trials Relevant to Musculoskeletal Conditions. *Arch Phys Med Rehabil*. 2018;99(1):129-136. doi:10.1016/j.apmr.2017.08.485
15. Alvarez G, Sola I, Sitjà-Rabert M, et al. A methodological review revealed that reporting of trials in manual therapy has not improved over time. *J Clin Epidemiol*. January 2020.
16. Cashin AG, Lee H, Bagg MK, et al. A systematic review highlights the need to improve the quality and applicability of trials of physical therapy interventions for low back pain. *J Clin Epidemiol*. 2020;126:106-115. doi:10.1016/j.jclinepi.2020.06.025
17. Bird L, Arthur A, Cox K. "Did the trial kill the intervention?" experiences from the development, implementation and evaluation of a complex intervention. *BMC Med Res Methodol*. 2011;11. doi:10.1186/1471-2288-11-24
18. Clark AM. What are the components of complex interventions in healthcare? Theorizing approaches to parts, powers and the whole intervention. *Soc Sci Med*. 2013;93:185-193. doi:10.1016/j.socscimed.2012.03.035
19. Cook CE, Thigpen CA. Five good reasons to be disappointed with randomized trials. *J Man Manip Ther*. 2019;27(2):63-65.
20. Boutron I, Tubach F, Giraudeau B, Ravaud P. Methodological Differences in Clinical Trials Evaluating Nonpharmacological and Pharmacological Treatments of Hip and Knee Osteoarthritis. *JAMA …*. 2003;290(8):1062-1070.
21. Boutron I, Guittet L, Estellat C, Moher D, Hróbjartsson A, Ravaud P. Reporting methods of blinding in randomized trials assessing nonpharmacological treatments. *PLoS Med*. 2007;4(2):0370-0380. doi:10.1371/journal.pmed.0040061
22. Hoffmann TC, Glasziou PP, Boutron I, et al. Better reporting of interventions: Template for intervention description and replication (TIDieR) checklist and guide. *BMJ*. 2014. doi:10.1136/bmj.g1687
23. Pillastrini P, Vanti C, Curti S, et al. Using PubMed search strings for efficient retrieval of manual therapy research literature. *J Manipulative Physiol Ther*. 2015. doi:10.1016/j.jmpt.2014.11.005

24. Barnett-Page E, Thomas J. Methods for the synthesis of qualitative research: A critical review. *BMC Med Res Methodol*. 2009. doi:10.1186/1471-2288-9-59
25. Guerrero-Bote VP, Moya-Anegón F. A further step forward in measuring journals' scientific prestige: The SJR2 indicator. *J Informetr*. 2012;6(4):674-688. doi:10.1016/j.joi.2012.07.001
26. Koes BW. How to evaluate manual therapy: Value and pitfalls of randomized clinical trials. *Man Ther*. 2004;9(4):183-184. doi:10.1016/j.math.2004.04.002
27. Rubinstein SM, Van Eekelen R, Oosterhuis T, De Boer MR, Ostelo RWJGJG, Van Tulder MW. The risk of bias and sample size of trials of spinal manipulative therapy for low back and neck pain: Analysis and recommendations. *J Manipulative Physiol Ther*. 2014;37(8):523-541. doi:10.1016/j.jmpt.2014.07.007
28. Lee PH, Tse ACY. The quality of the reported sample size calculations in randomized controlled trials indexed in PubMed. *Eur J Intern Med*. 2017;40:16-21.
29. Castellini G, Gianola S, Bonovas S, Moja L. Improving Power and Sample Size Calculation in Rehabilitation Trial Reports: A Methodological Assessment. *Arch Phys Med Rehabil*. 2016;97(7):1195-1201.
30. Schulz KF, Grimes DA. Sample size calculations in randomised trials: mandatory and mystical. *Lancet*. 2005;365(9467):1348-1353.
31. Van Tulder MW, Suttorp M, Morton S, Bouter LM, Shekelle P. Empirical Evidence of an Association Between Internal Validity and Effect Size in Randomized Controlled Trials of Low-Back Pain. *Spine (Phila Pa 1976)*. 2009;34(16):1685-1692.
32. Fregni F, Imamura M, Chien HF, et al. Challenges and recommendations for placebo controls in randomized trials in physical and rehabilitation medicine: A report of the international placebo symposium working group. *Am J Phys Med Rehabil*. 2010. doi:10.1097/PHM.0b013e3181bc0bbd
33. Chaibi A, Šaltytė Benth J, Bjørn Russell M. Validation of Placebo in a Manual Therapy Randomized Controlled Trial. *Sci Rep*. 2015;5(1):11774. doi:10.1038/srep11774
34. Cerritelli F, Verzella M, Cicchitti L, D'Alessandro G, Vanacore N. The paradox of sham therapy and placebo effect in osteopathy A systematic review. *Med (United States)*. 2016;95(35). doi:10.1097/MD.0000000000004728
35. Bialosky JE, Bishop MD, Penza CW. Placebo Mechanisms of Manual Therapy: A Sheep in Wolf's Clothing? *J Orthop Sport Phys Ther*. 2017;47(5):301-304. doi:10.2519/jospt.2017.0604
36. Bialosky JE, Bishop MD, Price DD, Robinson ME, George SZ. The mechanisms of manual therapy in the treatment of musculoskeletal pain:

A comprehensive model. *Man Ther*. 2009. doi:10.1016/j.math.2008.09.001
37. Bialosky JE, Beneciuk JM, Bishop MD, et al. Unraveling the Mechanisms of Manual Therapy: Modeling an Approach. *J Orthop Sports Phys Ther*. 2018;48(1):8-18.
38. Hancock MJ, Maher CG, Latimer J, McAuley JH. Selecting an appropriate placebo for a trial of spinal manipulative therapy. *Aust J Physiother*. 2006;52(2):135-138.
39. Hart T, Bagiella E. Design and Implementation of Clinical Trials in Rehabilitation Research. *Arch Phys Med Rehabil*. 2012;93(8, Supplement):S117--S126.
40. Benedetti F. Placebo effects: From the neurobiological paradigm to translational implications. *Neuron*. 2014;84(3):623-637. doi:10.1016/j.neuron.2014.10.023
41. Howick J, Bishop F, Collins G, et al. TIDieR—Placebo checklist explanation and elaboration. TIDieR—Placebo checklist explanation and elaboration. https://www.equator-network.org/library/reporting-guidelines-under-development/reporting-guidelines-under-development-for-clinical-trials/#TIDIERPLA. Accessed February 4, 2020.
42. Villamar MF, Contreras VS, Kuntz RE, Fregni F. The reporting of blinding in physical medicine and rehabilitation randomized controlled trials: A systematic review. *J Rehabil Med*. 2013;45(1):6-13. doi:10.2340/16501977-1071
43. Moustgaard H, Clayton GL, Jones HE, et al. Impact of blinding on estimated treatment effects in randomised clinical trials: meta-epidemiological study. *Bmj*. 2020;368:l6802--13.
44. Drucker AM, Chan A-W. Blindsided: challenging the dogma of masking in clinical trials. *Bmj*. 2020;368:m229. doi:10.1136/bmj.m229
45. Coronado RA, Bialosky JE. Manual physical therapy for chronic pain: the complex whole is greater than the sum of its parts. *J Man Manip Ther*. 2017;25(3):115-117. doi:10.1080/10669817.2017.1309344
46. Rossettini G, Carlino E, Testa M. Clinical relevance of contextual factors as triggers of placebo and nocebo effects in musculoskeletal pain. *BMC Musculoskelet Disord*. 2018;19(1):1-15. doi:10.1186/s12891-018-1943-8
47. Geri T, Viceconti A, Minacci M, Testa M, Rossettini G. Manual therapy: Exploiting the role of human touch. *Musculoskelet Sci Pract*. 2019;44:1-20. doi:10.1016/j.msksp.2019.07.008
48. Hrobjartsson A, Forfang E, Haahr MT, Als-Nielsen B, Brorson S. Blinded trials taken to the test: an analysis of randomized clinical trials that report tests for the success of blinding. *Int J Epidemiol*. 2007;36(3):654-663.
49. Sackett DL. Commentary: Measuring the success of blinding in RCTs: Don't, must, can't or needn't? *Int J Epidemiol*. 2007;36(3):664-665. doi:10.1093/ije/dym088
50. Sackett DL. Clinician-trialist rounds: 6. Testing for blindness at the end of

your trial is a mug's game. *Clin Trials*. 2011;8(5):674-676. doi:10.1177/1740774511419685
51. Hróbjartsson A, Boutron I. Blinding in randomized clinical trials: Imposed impartiality. *Clin Pharmacol Ther*. 2011;90(5):732-736. doi:10.1038/clpt.2011.207
52. Chang K-L, Fillingim R, Hurley RW, Schmidt S. Chronic pain management: nonpharmacological therapies for chronic pain. *FP Essent*. 2015.
53. Foster NE, Dziedzic KS, van der Windt DAWM, Fritz JM, Hay EM. Research priorities for non-pharmacological therapies for common musculoskeletal problems: nationally and internationally agreed recommendations. *BMC Musculoskelet Disord*. 2009;10:3. doi:10.1186/1471-2474-10-3
54. Dieppe P. Complex interventions. *Musculoskeletal Care*. 2004;2(3):180-186. doi:10.1002/msc.69
55. Lowe DB, Taylor MJ, Hill SJ. Cross-sectional examination of musculoskeletal conditions and multimorbidity: influence of different thresholds and definitions on prevalence and association estimates. *BMC Res Notes*. 2017;10(1):1-13. doi:10.1186/s13104-017-2376-4
56. Lowe DB, Taylor MJ, Hill SJ. Associations between multimorbidity and additional burden for working-age adults with specific forms of musculoskeletal conditions: a cross-sectional study. *BMC Musculoskelet Disord*. 2017;18(1):1-14. doi:10.1186/s12891-017-1496-2
57. Duffield SJ, Ellis BM, Goodson N, et al. The contribution of musculoskeletal disorders in multimorbidity: Implications for practice and policy. *Best Pract Res Clin Rheumatol*. 2017;31(2):129-144.
58. Treweek S, Zwarenstein M. Making trials matter: pragmatic and explanatory trials and the problem of applicability. *Trials*. 2009;10:37. doi:10.1186/1745-6215-10-37
59. Negrini S, Arienti C, Pollet J, et al. Clinical replicability of rehabilitation interventions in randomized controlled trials reported in main journals is inadequate. *J Clin Epidemiol*. 2019;114:108-117.
60. Yamato TP, Maher CG, Saragiotto BT, Hoffmann TC, Moseley AM. How completely are physiotherapy interventions described in reports of randomised trials? *Physiotherapy*. 2016;102(2):121-126.
61. Boutron I, Altman DG, Moher D, et al. CONSORT Statement for Randomized Trials of Nonpharmacologic Treatments: A 2017 Update and a CONSORT Extension for Nonpharmacologic Trial Abstracts. *Ann Intern Med*. 2017;167(1):40-47. doi:10.7326/M17-0046
62. Groeneweg R, Rubinstein SM, Oostendorp RAB, Ostelo RWJG, Van Tulder MW. Guideline for Reporting Interventions on Spinal Manipulative Therapy: Consensus on Interventions Reporting Criteria List for Spinal Manipulative Therapy (CIRCLe SMT). *J Manip Physiol Ther*. 2017;40(2):61-70.
63. Page P, Hoogenboom B, Voight M. Improving the Reporting of

Therapeutic Exercise Interventions in Rehabilitation Research. *Int J Sports Phys Ther*. 2017;12(2):297-304. http://www.ncbi.nlm.nih.gov/pubmed/28515984%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5380872.

64. Hefford C, Abbott JH, Baxter GD, Arnold R. Outcome measurement in clinical practice: practical and theoretical issues for health related quality of life (HRQOL) questionnaires. *Phys Ther Rev*. 2013;16(3):155-167.
65. Gagnier JJ, Moher D, Boon H, Beyene J, Bombardier C. Investigating clinical heterogeneity in systematic reviews: a methodologic review of guidance in the literature. *BMC Med Res Methodol*. 2012;12(1):1.
66. Smith T. The challenges of heterogeneity across physiotherapy trials when considering meta-analysis. *Phys Ther Rev*. 2013;19(1):39-40.
67. Kyte DG, Calvert M, van der Wees PJ, ten Hove R, Tolan S, Hill JC. An introduction to patient-reported outcome measures (PROMs) in physiotherapy. *Physiotherapy*. 2015;101(2):119-125.
68. Calvert M, Blazeby J, Altman DG, Revicki DA, Moher D, Brundage MD. Reporting of patient-reported outcomes in randomized trials: The CONSORT PRO extension. *JAMA - J Am Med Assoc*. 2013. doi:10.1001/jama.2013.879

Table 1: Type of limitations and reasons reported by authors

Category 1 - **Randomization and/or Concealment**	
Unbalanced groups	*After randomization, the number of subjects allocated to each group was unequal.*
Lack of stratification	*Randomization method did not include stratification.*
Selection Bias not excluded	*Randomization methods were not the most recommended (e.g., ballot boxes) or allocation was not concealed.*
Category 2 - Blinding	
Statistician blinding	*People performing the statistical analysis were not blinded.*
Outcome assessor blinding	*People in charge of data collection were not blinded.*
Patient blinding	*Patients were not blinded with respect to the study groups.*
Care provider blinding	*People providing the interventions were not blinded with respect to the study groups.*
Category 3 - **Lost to Follow-up / Dropouts**	
Unbalanced dropouts	*Dropout rate between study groups was unequal.*
Many lost to follow-up	*Discontinuation rate was superior than expected/estimated leading to underpowered studies and less precise results.*
Category 4 - **Sample size**	
Unpowered study	*Inability to detect differences between groups due to sample size.*
Convenience sampling	*The sampling method was linked to specific study needs.*
Recruitment less than expected	*The sample size did not reach the estimated number due to recruitment difficulties.*
Category 5 - **Inadequate control**	
Active control	*An active intervention (e.g., exercise or usual care) was selected as control.*
Simulation (Sham) not optimal	*Some effects related to sham interventions could not be ruled out.*
Different parameters between groups	*The dosage, the number of sessions or their duration was different in the control group.*
Different baseline parameters.	*Comparison of baseline data showed significant differences in the control group*
Category 6 - **Funding**	

Limited funding	*The limited or lack of funding affected the study progress or completion.*
Category 7 - **Type of setting / care provider/s**	
Unicentric	*Study was conducted recruiting participants from a single center.*
Number of care providers	*Intervention delivered by a single investigator.*
Category 8 - **Inadequate subjects**	
Asymptomatic subjects	*Leading to a limited clinical interpretation of the results.*
Diagnostic criteria	*Lack of standardized diagnostic criteria to include participants.*
Very specific population	*Inclusion criteria considered too restricted. (e.g., single gender, athletes only, education level or race).*
Category 9 - **Study length and follow-up**	
No long term follow up	*Only short-term effect was evaluated.*
Missing data due to study length	*Dropouts or losses to follow-up derived from a very long follow-up time.*
Unequal follow-up	*Follow-up between the study groups was different.*
Category 10 - **Intervention**	
Additive effects	*It was not possible to know the net effect of every component in multimodal treatments.*
Limited treatment area	*The intervention was applied to a very specific body region (e.g., single joint or vertebral segment).*
Treatment parameters not standardized:	*Specific parameters of manual techniques were not standardized (e.g., amount of force, direction or speed).*
Lack of individualization	*A treatment protocol was applied without adapting the interventions to the characteristics of the individuals or to the diagnostic criteria.*
Therapist profile	*Care providers required specific training on the intervention or care providers had different previous experience*
Category 11 - **Outcome measures**	
Reliability of the measurement	*Lack or low reliability of outcome measures (e.g., self-reported results).*
Lack of measurement data	*Some relevant data was not collected during the study (e.g., demographics or clinical variables that would potentially provide interesting findings).*
Measurement sensitivity	*Outcome measures were not sensitive enough to detect subtle changes (e.g., use of ordinal scales).*
Validity of the measurement	*The selected assessment instrument was not originally validated for that specific population.*
Suboptimal statistical analysis:	*Findings were not adjusted for covariates.*

Category 12 - **Compromised generalization of the results** *(This category includes those self-acknowledged limitations that were mentioned specifically in regards to their impact on the generalization of the results or external validity.)*	
Compromised generalization due to type of setting.	*The characteristics of the setting where the study was conducted limited the generalization of the results*
Compromised generalization due to specific population	*The subjects' characteristics of the sample limited the generalization of the results.*
Compromised generalization due to measurement	*The measurement instrument selected to assess the effect of the intervention sample limited the generalization of the results.*
Compromised generalization due to intervention	*The type of intervention or the way it was delivered limited the generalization of the results.*

Table 2: Frequency and type of reported limitations

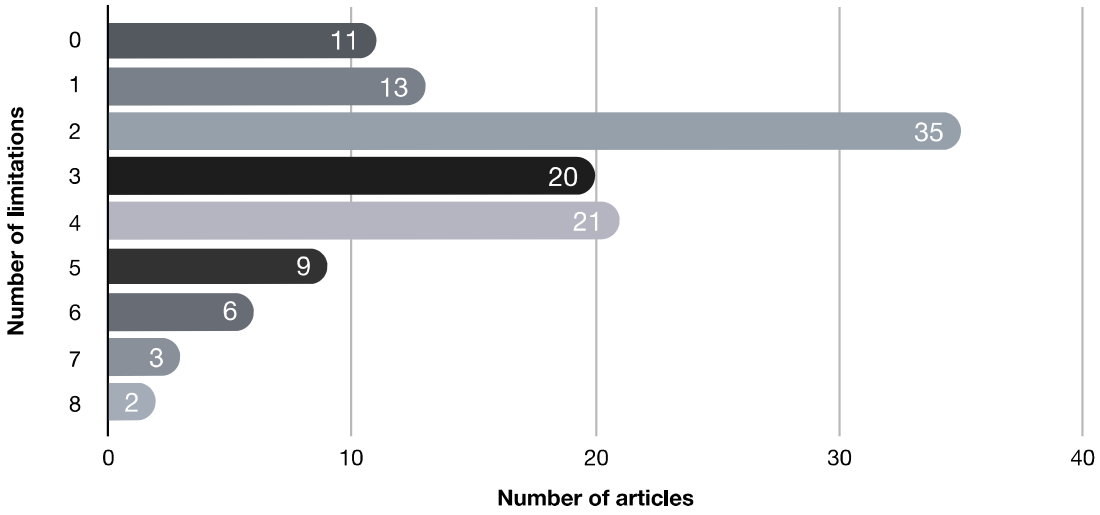
Randomization and/or Concealment		**N articles**	**Frequency [CI 95%]**
N=13 / 10.8% [5.9 to 17.8]	Unbalanced groups	2	15.4% [1.9 to 45.4]
	Lack of stratification	3	23.1% [5.1 to 53.8]
	Selection bias not excluded	8	61.5% [31.5 to 86.1]
Blinding		**N articles**	**Frequency [CI 95%]**
N=34 / 28.3% [20.5 to 37.3]	Statistician blinding	1	2.9% [0.1 to 15.3]
	Outcome assessor blinding	4	11.8% [3.3 to 27.5]
	Patient blinding	24	70.6% [52.5 to 84.9]
	Care provider blinding	17	50.0% [32.4 to 67.6]
Lost to Follow-up / Dropouts		**N articles**	**Frequency [CI 95%]**
N=18 / 15% [9.1 to 22.7]	Unbalanced dropouts	1	5.6% [0.1 to 2.67]
	Many lost to follow-up	17	94.4% [72.7 to 99.9]
Sample size		**N articles**	**Frequency [CI 95%]**
N=57 / 47.5% [38.3 to 56.8]	Unpowered study	50	87.7% [76.3 to 94.9]
	Convenience sampling	2	3.5% [0.4 to 12.1]
	Recruitment less than expected	5	8.8 [2.9 to 19.3]
Inadequate control		**N articles**	**Frequency [CI 95%]**
N=39 / 32.5% [24.2 to 41.7]	Active control.	14	35.9% [21.2 to 52.8]
	Simulation (Sham) not optimal	16	41% [25.5 to 57.9]
	Different parameters between groups	4	10.3% [2.9 to 24.2]
	Different baseline parameters	5	12.8% [4.3 to 27.4]
Funding		**N articles**	**Frequency [CI 95%]**
N=5 / 4.2% [1.4 to 9.5]	Limited funding	5	100% [47.8 to 100]
Type of setting / care providers		**N articles**	**Frequency [CI 95%]**
N=9 / 7.5% [3.5 to 13.8]	Unicentric	8	88.9% [51.7 to 99.7]
	Number of care providers	1	11.1% [0.3 to 48.2]
Inadequate subjects		**N articles**	**Frequency [CI 95%]**
N=31 / 25.8% [18.3 to 34.6]	Asymptomatic subjects	4	12.9% [3.6 to 29.8]
	Diagnostic criteria	5	16.1% [5.4 to 33.7]
	Very specific population	23	74.2% [55.4 to 88.1]
Study length and follow-up		**N articles**	**Frequency [CI 95%]**
N=40 / 33.3% [25.o to 42.5]	No long term follow up	35	87.5% [73.1 to 95.8]
	Missing data due to study length	2	5% [0.6 to 16.9]

	Unequal follow-up	3	7.5% [1.5 to 20.4]
Intervention		**N articles**	**Frequency [CI 95%]**
N=33 / 27.5% [19.7 to 36.4]	Additive effects Limited treatment area Treatment parameters not standardized Lack of individualization Therapist profile	6 8 17 5 1	18.2% [6.9 to 35.5] 24.2% [11.1 to 42.3] 51.5% [33.5 to 69.2] 15.2% [5.1 to 31.9] 3% [0.1 to 15.8]
Outcome measures		**N articles**	**Frequency [CI 95%]**
N=35 / 29.2% [21.2 to 38.2]	Reliability of the measurement Lack of measurement data Measurement sensitivity Validity of the measurement Suboptimal statistical analysis	9 12 2 11 4	25.7% [12.5 to 43.3] 34.3% [19.1 to 52.2] 5.7% [0.7 to 19.2] 31.4% [16.9 to 49.3] 11.4% [3.2 to 26.7]
Compromised generalization (CG) of the results (n=26)		**N articles**	**Frequency [CI 95%]**
N=26 / 21.7% [14.7 to 30.1]	CG due to type of setting CG due to specific population CG due to measurement CG due to intervention	3 16 2 5	11.5% [2.4 to 30.2] 61.5% [40.5 to 79.8] 7.7% [0.9 to 25.1] 19.2% [6.5 to 39.4]

95% confidence interval (CI)

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

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



HIGHLIGHTS

What is new?

Key Findings

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

What this adds to what was known?

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

What is the implication and what should change now?

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

AUTHOR STATEMENT

GA conceptualized and designed the study. GA and GU were responsible to create the first version of the data extraction form and GA, IS, MS, AF, CF and GU piloted the form and contributed to reach the final version. IS was responsible of the search strategy and articles retrieval. IS, MS, AF, CF and GU were involved in the data extraction process forming pairs with GA. RN and GA were responsible of data analysis. GA provide the first manuscript draft which received critical revision by GU, IS and XB. All authors read and approved the final manuscript.

The present study is part of the PhD thesis of Gerard Alvarez. If the manuscript is deemed appropriate for his publication and as a requirement of the university I would appreciate that the following statement could be included in the author's information:

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

The author also wants to state that, before submission, the present manuscript has been revised by a professional editing service to ensure quality and adherence to JCE standards

Conflict of Interest

The authors declare that they have no competing interests.