Improving Power and Sample Size Calculation in Rehabilitation Trial Reports: A Methodological Assessment

Greta Castellini, PT, MSc, Silvia Gianola, PT, OMT, MSc, Stefanos Bonovas, MD, MSc, PhD, Lorenzo Moja, MD, MSc, PhD

PII: S0003-9993(16)00152-0

DOI: 10.1016/j.apmr.2016.02.013

Reference: YAPMR 56468

To appear in: *ARCHIVES OF PHYSICAL MEDICINE AND REHABILITATION*

Received Date: 5 November 2015

Revised Date: 16 February 2016

Accepted Date: 16 February 2016

Please cite this article as: Castellini G, Gianola S, Bonovas S, Moja L, Improving Power and Sample Size Calculation in Rehabilitation Trial Reports: A Methodological Assessment, *ARCHIVES OF PHYSICAL MEDICINE AND REHABILITATION* (2016), doi: 10.1016/j.apmr.2016.02.013.

Title: IMPROVING POWER AND SAMPLE SIZE CALCULATION IN REHABILITATION TRIAL REPORTS: A METHODOLOGICAL ASSESSMENT

Contact author:

Greta Castellini, PT, MSc

1 Department of Biomedical Sciences for Health, University of Milan, Italy

2 Unit of Clinical Epidemiology, I.R.C.C.S. Orthopedic Institute Galeazzi, Milan, Italy

+39 3491242504

gre.caste@gmail.com

Pending authors:

Silvia Gianola, PT, OMT, MSc

1 Center of Biostatistics for Clinical Epidemiology, Department of Health Science, University of Milano-Bicocca, Monza, Italy

2 Unit of Clinical Epidemiology, I.R.C.C.S. Orthopedic Institute Galeazzi, Milan

silvia.gianola@gmail.com

Stefanos Bonovas, MD, MSc, PhD

1 Humanitas Clinical and Research Center, Milan, Italy

sbonovas@gmail.com

Lorenzo Moja, MD, MSc, PhD

1 Department of Biomedical Sciences for Health, University of Milan, Italy

2 Unit of Clinical Epidemiology, I.R.C.C.S. Orthopedic Institute Galeazzi, Milan, Italy

lorenzo.moja@unimi.it

Conflict of interest: None.

IMPROVING POWER AND SAMPLE SIZE CALCULATION IN REHABILITATION TRIAL REPORTS: A METHODOLOGICAL ASSESSMENT

ABSTRACT

Objective

To systematically assess the reporting of sample size calculation in RCTs on rehabilitation interventions for mechanical low-back pain (mLBP).

Study selection

We conducted an electronic database search for RCTs published from 1968 through February 2015 and included in Cochrane Systematic Reviews (SRs).

Data extraction

Two investigators independently applied an ad hoc six-item checklist derived from the CONSORT 2010 statement recommendations to extract data on sample size calculation. Primary outcome was the proportion of RCTs that reported sample size calculation; secondary outcome was the completeness of sample size analysis reporting. We also evaluated reporting' improvement over time.

Data synthesis

Sample size calculation was reported in 80 (36.0%) of the 222 eligible RCTs included in 14 Cochrane SRs. Only 13 (16.3%) of these RCT reports gave a complete description and about half reported four or more of the six elements of sample size calculation (median=4, IQR 3–5). Completeness of reporting sample size calculation improved from 1968 to 2013; beginning in 2005, the number of RCT reports containing this information increased over those not reporting it.

Conclusions

Despite improvement, reporting of sample size calculation and power analysis remains inadequate, limiting the reader's ability to assess the quality and accuracy of rehabilitation studies.

Keywords: rehabilitation, power, sample size calculation, randomized clinical trial, design

Abbreviations

RCT, randomized controlled trial

CONSORT, Consolidated Standards of Reporting Trials

1. INTRODUCTION

Well-designed, properly executed RCTs provide the most reliable evidence on the effectiveness of health care interventions [1]. The validity of an RCT depends on several key factors that should be adequately reported: the sample size calculation is one of them. Sample size is related to statistical power, which derives from beta error or type II error [2,3]: it represents the likelihood of failure to reject the null hypothesis when, in fact, it should be rejected. The investigator's aim is to minimize this type of error by increasing the sample size. Sample size calculation is essential in study design because a low-powered study may fail to yield significant results and detect relevant clinical effects. Its description is fundamental in any published report so that readers can base their assessment on what is reported rather than rely on assumptions about how the study authors arrived at their results. However, sample size calculation is not always adequately reported [4-6].

In order to ensure quality in trial conduction, the Consolidated Standards of Reporting Trials (CONSORT) 2010 statement recommends that authors provide a clear description of sample size calculation methods and assumptions as follows: the estimated outcomes in each group (minimum important treatment effect or effect size), the level of significance (alpha or type I error), the statistical power (beta or type II error), and, for continuous outcomes, the assumed standard deviation of the measurements [4,7,8]. In addition, the CONSORT guidelines also recommend reporting the primary outcome on which important differences between two groups are determined. Authors should therefore decide and state a priori the fixed values for parameter assumptions. Although the number of reports of RCTs in rehabilitation has been increasing [9], the majority of studies are based on clinical observations with small sample sizes and inadequate reporting of essential information [10].

The purpose of the present review is to systematically assess the quality of reporting of power and sample size calculation in RCTs comparing mechanical low-back pain rehabilitation interventions and included in Cochrane systematic reviews.

2. METHODS

2.1 Search strategy and study selection

We conducted an electronic database search for systematic reviews published between 1968 and February 2015 limited to The Cochrane Database for Systematic Reviews. Search terms 'back pain' and 'rehabilitation' were run in "title,abstract,keywords" search tab in advance search strategy. We included a systematic review if the title or the abstract presented mechanical low-back pain as the disease target and the intervention was rehabilitative, as defined by the National Library of Medicine [11]. We did not take into account interventions other than therapeutic rehabilitation (e.g., prevention) or involving population subgroups (e.g., pregnancy). From the eligible systematic reviews, we extracted all included trials with a randomized study design and published in English, Italian, Spanish or French. After removing duplicates of RCTs, two researchers (GC, SG) independently screened the title and abstract of all potentially eligible RCTs. Disagreements were resolved by consensus.

2.2 Data Extraction

We extracted the general characteristics of RCTs: year of publication, number of authors, first author's geographic region (Europe, North and South America, Asia and Australia), journal that published the study, and funding source. We developed an ad hoc checklist derived from the CONSORT checklist to extract data on sample size calculation. The checklist was upload on Distiller SR, a web-based database for data management.

We examined whether the RCT report included a power analysis in the Methods section and, if so, whether the description of the sample size calculation was CONSORT-compliant. Following the CONSORT checklist [7], we assessed the description for reporting of six sample size calculation components: (1) type I error, or alpha, (2) type II error, beta, or power, (3) assumption of expected treatment effect of the intervention (i.e., the difference between group means as effect size or minimal important difference and relative risk), and (4) the assumed variability expressed as a standard deviation or a variance or an intraclass correlation coefficient. We also looked for (5) the outcome on

which sample size calculation was based, and (6) whether there was an adjustment to accommodate attrition rate. In addition, we extracted from the Methods section the sample size planned (i.e. as resulted from the sample size calculation procedure) and from the Results section the actual number of participants randomized (N) according to the CONSORT flow diagram. If there was no statement or CONSORT flow diagram reporting the number of patients randomized, we extracted it from implicit information (i.e., “enrolled” or “included”). When articles reported the sample size calculation, we examined whether there was a discrepancy between the planned sample size and the number of participants randomized. Moreover, we asked whether sample size reporting might be influenced by the funding status of the RCT.

Data extraction was independently performed by two reviewers (GC, SG). Disagreements were reconciled via consensus.

2.3 Statistical Methods

Descriptive statistics are presented as medians and interquartile ranges (IQR), or percentages when appropriate. The non-parametric matched-pairs Wilcoxon signed-rank test, and the Chi-squared test, were used for the statistical evaluations. For hypothesis testing, a probability level lower than 0.05 was considered to be statistically significant. All statistical tests were two-sided. Stata software was used for all statistical analyses (Stata Corp., College Station, TX, USA).

3. RESULTS

3.1 Study selection

We identified 14 relevant Cochrane systematic reviews in the Cochrane Library [12-25]. Sixty out of 301 RCTs included in these 14 systematic reviews were excluded because they were duplicates or multiple publications of the same RCT, 7 were excluded as their full text could not be retrieved, and 12 were

excluded because they did not satisfy the language criterion. A final total of 222 RCTs was included in our review. ***Figure 1***.

3.2 General characteristics

The 222 eligible RCT reports were published in 78 journals. Most were published in *Spine* (22.5%, n=50), followed by *Journal of Manipulative and Physiological Therapeutics* (4.5%, n=10) *Pain*, *British Medical Journal*, and *Archives of Physical Medicine and Rehabilitation* (4.1%, n=9), and *Clinical Journal of Pain* (3.6%, n=8).

Some 32 countries were indicated as the country of publication, with the three top countries being the United States (18.9%, n=42), the United Kingdom (13.1%, n=29) and the Netherlands (9.9%, n=22); most studies were published (59.5%, n=132) by European researchers. The period of RCTs publication was from 1968 to 2013. The characteristics of the RCTs are reported in ***Table 1***.

3.3 Sample size calculation

3.3.1 Reporting

Only 80 (36.0%) of the 222 RCTs reported sample size calculation. However, there was a significant improvement of sample size calculation reporting over time ***Figure 2***. We found that 13.3% (11 of 83) of trials published on or before 1996 reported sample size calculation compared to 49.6% (69 of 139) of trials published on or after 1997 (Chi-squared=29.85, d.f.=1, p<0.001). Furthermore, we found an association between reporting of a funding source and sample size calculation reporting. In particular, 48.8% (61 of 125) of the trials reporting a funding source were also reporting a sample size calculation compared to only 19.6% (19 of 97) of the trials not reporting a funding source (Chi-squared=20.22, d.f.=1, p<0.001). This association was very strong in the post-CONSORT era with 61.4% (54 of 88) of the trials reporting a funding source also reporting a sample size calculation vs. 29.4% (15 of 51) of the RCTs not reporting a funding source (Chi-squared=13.19, d.f.=1, p<0.001). However, it was not significant in the pre-CONSORT era (18.9% vs. 8.7%, Chi-squared=1.86, d.f.=1, p=0.17); but data were scarce.

144

145 ***3.3.2 Complete description of sample size calculation***

146 Thirteen (16.3%) of the 80 RCTs reporting sample size calculation gave an adequate description of
147 the a priori sample size calculation, with all six elements provided in compliance with CONSORT
148 guidelines. Half of the RCTs reported at least four out of six elements. ***Figure 3***.
149 Of the six CONSORT components required for sample size calculation, the three most frequently
150 reported were the power (91.3%, n=73), followed by the assumption concerning the expected
151 treatment effect of the intervention (86.3%, n=69), and the alpha error or type I error (85.0%, n=68).
152 Adjustment to accommodate attrition was the least frequently reported element (32.5%, n=26).

153

154 ***3.3.3 Characteristics of each element reported***

155 Each element could be expressed in a different way; common expressions for elements are presented
156 in ***Table 2***. Power was usually defined as $1 - \beta$ (82.5%, n=66). The minimal important difference
157 (MID) was the assumed value for the detection of treatment effect most often reported in the 80 trials
158 (46.3%, n=37). Concerning the outcome on which the calculation was based, all RCTs evaluated
159 continuous outcomes: disability was the one most often reported (42.5%, n=34), followed by pain
160 (22.5%, n=18).

161

162 **3.4 Discrepancy between planned and randomized sample size**

163 Planned sample size was reported in 72 out of 80 RCTs. In the remaining 8 RCTs (10.0%) that
164 reported the sample size calculation, the planned number of participants was not stated. The median
165 number of participants needed to prove sufficient power was 120 (range: 17–2000), whereas the
166 median of the number of participants randomized among these 72 RCTs was 133 (range: 21–741).
167 The number of participants randomized was lower than the number of those planned in 17 RCTs
168 (23.6%), equal in 13 (18.0%), and higher in 42 (58.4%); ***Figure 4*** showed the discrepancy between
169 sample size planned and the number of randomized participants when the number obtained by the
170 sample size calculation increased.

4. DISCUSSION

Reporting of sample size calculation in RCTs on low-back pain rehabilitation is often incomplete. We found that numerous RCTs published between the 1960s and the present failed to report a priori sample size calculation, barring readers from understanding whether calculation was done and whether done correctly. Among the RCTs reporting a priori sample size calculation, only a minority gave a complete description of the elements used. Nevertheless, the reporting of sample size calculation and its components has increased over years; since 2005 more RCTs report sample size calculation than those that do not. Moreover, our results showed that the publication of the CONSORT statement has increased authors' awareness of high quality reporting compared to the pre-CONSORT era. Despite this, assessing the quality of the reporting does not necessary reflect the quality of the underlying research: it is fundamental distinguishing between 'what researchers do' and 'what researchers report'. For instance, the assessment of risk of bias in a RCT arises ambiguity between the quality of reporting and the quality of the research [26].

Our findings are consistent with a previous review of the general medical literature that described poor compliance by authors with CONSORT guidelines. Similarly, a review of physical medicine and rehabilitation trials published between 1998 and 2008 found that reporting had improved somewhat, with only slightly more than half of the articles (57.3%) published in 2008 reporting sample size analysis [10].

Conducting responsible research entails complete, accurate reporting in a transparent fashion according to international guidelines. To ensure high quality in conducting a clinical trial, it is not sufficient to state the sample size without giving a description of how it was calculated. More than half of the RCTs with a priori sample size analysis included in our review reported fewer than four of the six elements required for replication of calculations. A recent review (ACTTION Systematic Review) found that half of the published analgesic clinical trials gave an incomplete description of sample size calculation [2].

Sample size calculation is usually based on a single outcome, chosen as a primary measure: specifying it helps researchers to clarify the initial basis upon which an RCT is built, besides simplifying interpretation, judgment, and use of findings [27]. We noted that more than half of the RCTs stated the primary endpoint, similar to the rates reported in a previous review in physical medicine trials [10]. In the literature, *disability* and *pain* are the most frequently investigated outcomes in low-back pain rehabilitation: several authors have recommended including these measurements in the back-specific core outcome sets because they are most relevant to patients, health care practitioners, regulators, industry representatives, and policy-makers [28]. They were also the elective outcome measures most often used in RCTs according to our and a recent review which found a low frequency of reporting outcome and intervention descriptions, reflecting a multidimensional lack of quality in rehabilitation RCTs [29].

Among the RCTs in which a power analysis was performed, 72 reported the planned sample size. In two out of three of these RCTs the randomized sample size was larger than that planned, and in a small proportion (30%) the randomized sample size was smaller than that planned. While authors are always encouraged to include more than the minimum number of participants to compensate for loss to follow-up, overrecruitment to account for attrition is unjustifiable both economically and ethically – economically unsound because of the high costs of clinical trials and ethically questionable because of potential harm to patients. Except for trials on rare diseases or early-phase trials, underpowered studies are unethical because they may fail to yield significant results, are more likely to be inconclusive and produce more false negatives [1,30,31]. However, trials with an overly large sample size may waste resources in terms of patients, time and funding. Authors should aim to achieve robust research findings by calculating an adequate sample size, using time and resources in the best cost-effective manner [32] and in collaboration with experienced biostatisticians and methodologist-researchers [33].

Our results show that funding status influences the quality of reporting. Building a sustainable funding scheme for clinical comparative research in areas less explored, i.e., the “orphan areas” such as anesthesiology or orthopedics, is critical to support evidence-based practice in medical research [34].

Funding is fundamental to obtaining more resources in terms of personnel and to make the research process more efficient. Economic support is important in both pharmacological research and research areas where public health needs are changing. For example, rehabilitation for low-back pain has increased its importance in both primary care, where rehabilitation as intervention plays a central role in LBP management, and research [9]; therefore, evidence-based rehabilitation has grown. When the aim is to translate results from research to practice, it is essential to focus on how the evidence is generated: the quality of RCTs can directly influence the conclusions of systematic reviews, with the risk that trials failing to detect a real difference between treatment effects may inflate the results of meta-analyses, obfuscating the decision-making process of physical therapists. RCT reports should provide essential information so that readers can make better decisions in clinical practice, especially in the rehabilitation of low-back pain, an increasingly common health problem with a substantial community and financial burden [35,36].

Future studies should assessed the quality of reporting of other essential elements for clinicians in rehabilitation. For instance, an adequate and satisfied description of the experimental intervention should be crucial, as well as the description of the target population and the outcomes selection. Maybe a multidimensional lack of reporting of information exists, reflecting difficulties in transferring the research's results in clinical practice.

4.1 Study Limitations

This study focused only on the reporting of sample size calculation and its components as described in the Methods section of RCTs. It would have been interesting to compare the final publication with the published protocol in order to explore whether the absence of some elements was limited to the research article or were included in the research protocol. This was not possible because our sample comprised a wide range of RCTs published from 1968 to 2013.

5. CONCLUSION

Sample size calculation is essential to demonstrate that a trial is adequately designed to detect a likely real effect or association, if such exists, in a given population [32]. Although some elements are difficult to define, the assumptions made in the calculation should be reported in a transparent fashion. The CONSORT statement provide a standard guidance for authors to prepare reports of trial findings and to facilitating their complete and transparent reporting. As well, the SPIRIT (Standard Protocol Items Recommendation for Interventional Trials) initiative recently has strengthened the purpose to improve transparency in the trial protocols [37]. Furthermore, Cook et al. have just created a more extensive set of elements for adequate reporting of this process in trial protocols and results, providing also justifications for sample size calculation' assumption [27]. Just as researchers should be encouraged to use these guidelines so, too, journal editors and peer reviewers should impose stricter criteria for adequate and transparent reporting. In addition, the sharing of software could help to simplify sample size calculation. Improving the methodological quality of RCTs, and all types of trials, will go some way to ensure the validity of results, reproducibility of research, and dissemination of results from research to practice.

Funding: none

BIBLIOGRAPHY

1. 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 : the journal of the American Medical Association.* Feb 27 2013;309(8):814-822.

2. McKeown A GJ, McDermott MP, Pawlowski JR, Poli JJ, Rothstein D, Farrar JT, Gilron I, Katz NP, Lin AH, Rappaport BA, Rowbotham MC, Turk DC, Dworkin RH, Smith SM. Reporting of Sample Size Calculations in Analgesic Clinical Trials: ACTTION Systematic Review. *The Journal of Pain.* 2015;16(3 (March)):199-206.

3. Schulz KF, Grimes DA. Sample size calculations in randomised trials: mandatory and mystical. *Lancet.* Apr 9-15 2005;365(9467):1348-1353.

4. Rutterford C, Taljaard M, Dixon S, Copas A, Eldridge S. Reporting and methodological quality of sample size calculations in cluster randomized trials could be improved: a review. *Journal of clinical epidemiology.* Dec 15 2014.

5. Koletsi D, Pandis N, Fleming PS. Sample size in orthodontic randomized controlled trials: are numbers justified? *European journal of orthodontics.* Feb 2014;36(1):67-73.

6. Ayeni O, Dickson L, Ignacy TA, Thoma A. A systematic review of power and sample size reporting in randomized controlled trials within plastic surgery. *Plastic and reconstructive surgery.* Jul 2012;130(1):78e-86e.

7. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ.* 2010;340:c869.

8. Antes G. The new CONSORT statement. *BMJ.* 2010;340:c1432.

9. Castellini G. GS, Banfi G., Bonovas S., Moja L. Mechanical low back pain: secular trend and intervention topics of randomized controlled trials. Physiotherapy Canada 2016; 68(1);61–63.

10. 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. *Archives of physical medicine and rehabilitation.* Feb 2011;92(2):306-315.

11. MeSH. http://www.ncbi.nlm.nih.gov/mesh. In: National Library Medicine controlled vocabulary NIoHN e, accessed in September 2013.

12. Rubinstein SM, Terwee CB, Assendelft WJ, de Boer MR, van Tulder MW. Spinal manipulative therapy for acute low back pain: an update of the cochrane review. *Spine.* Feb 1 2013;38(3):E158-177.

13. Yousefi-Nooraie R, Schonstein E, Heidari K, et al. Low level laser therapy for nonspecific low-back pain. *The Cochrane database of systematic reviews.* 2008(2):CD005107.

14. Walker BF, French SD, Grant W, Green S. Combined chiropractic interventions for low-back pain. *The Cochrane database of systematic reviews.* 2010(4):CD005427.

15. Hayden JA, van Tulder MW, Malmivaara A, Koes BW. Exercise therapy for treatment of non-specific low back pain. *The Cochrane database of systematic reviews.* 2005(3):CD000335.

16. Heymans MW, van Tulder MW, Esmail R, Bombardier C, Koes BW. Back schools for non-specific low-back pain. *The Cochrane database of systematic reviews.* 2004(4):CD000261.

17. Furlan AD, Imamura M, Dryden T, Irvin E. Massage for low-back pain. *The Cochrane database of systematic reviews.* 2008(4):CD001929.

18. Clarke JA, van Tulder MW, Blomberg SE, et al. Traction for low-back pain with or without sciatica. *The Cochrane database of systematic reviews.* 2007(2):CD003010.

19. Khadilkar A, Odebiyi DO, Brosseau L, Wells GA. Transcutaneous electrical nerve stimulation (TENS) versus placebo for chronic low-back pain. *The Cochrane database of systematic reviews.* 2008(4):CD003008.

20. Urrutia G, Burton AK, Morral A, Bonfill X, Zanoli G. Neuroreflexotherapy for non-specific low-back pain. *The Cochrane database of systematic reviews.* 2004(2):CD003009.

21. Rubinstein SM, van Middelkoop M, Assendelft WJ, de Boer MR, van Tulder MW. Spinal manipulative therapy for chronic low-back pain. *The Cochrane database of systematic reviews.* 2011(2):CD008112.

22. Henschke N, Ostelo RW, van Tulder MW, et al. Behavioural treatment for chronic low-back pain. *The Cochrane database of systematic reviews.* 2010(7):CD002014.

23. Ebadi S, Henschke N, Nakhostin Ansari N, Fallah E, van Tulder MW. Therapeutic ultrasound for chronic low-back pain. *The Cochrane database of systematic reviews.* 2014;3:CD009169.

24. Kamper SJ, Apeldoorn AT, Chiarotto A, et al. Multidisciplinary biopsychosocial rehabilitation for chronic low back pain. *The Cochrane database of systematic reviews.* 2014;9:CD000963.

25. Wegner I, Widyahening IS, van Tulder MW, et al. Traction for low-back pain with or without sciatica. *The Cochrane database of systematic reviews.* 2013;8:CD003010.

26. Higgins JPT, Green S. Chapter 8: Assessing risk of bias in included studies. *Cochrane Handbook for Systematic Reviews of Interventions: Version 5.0.1*: The Cochrane Collaboration; 2008.

27. Cook JA, Hislop J, Altman DG, et al. Specifying the target difference in the primary outcome for a randomised controlled trial: guidance for researchers. *Trials.* 2015;16(1):12.

28. Froud R, Patterson S, Eldridge S, et al. A systematic review and meta-synthesis of the impact of low back pain on people's lives. *BMC musculoskeletal disorders.* 2014;15:50.

29. Gianola S, Castellini G, Agostini M, et al. Reporting of Rehabilitation Intervention for Low Back Pain in Randomized Controlled Trials: Is the Treatment Fully Replicable? Spine November 2015 - Ahead of Print.

30. Maggard MA, O'Connell JB, Liu JH, Etzioni DA, Ko CY. Sample size calculations in surgery: are they done correctly? *Surgery.* Aug 2003;134(2):275-279.

31. Charles P, Giraudeau B, Dechartres A, Baron G, Ravaud P. Reporting of sample size calculation in randomised controlled trials: review. *BMJ.* 2009;338:b1732.

32. Fitzner K, Heckinger E. Sample size calculation and power analysis: a quick review. *The Diabetes educator.* Sep-Oct 2010;36(5):701-707.

33. Ioannidis JP, Greenland S, Hlatky MA, et al. Increasing value and reducing waste in research design, conduct, and analysis. *Lancet.* Jan 11 2014;383(9912):166-175.

34. Feasibility and challenges of independent research on drugs: the Italian medicines agency (AIFA) experience. *European journal of clinical investigation.* Jan 2010;40(1):69-86.

35. March L, Smith EU, Hoy DG, et al. Burden of disability due to musculoskeletal (MSK) disorders. *Best practice & research. Clinical rheumatology.* Jun 2014;28(3):353-366.

36. Maniadakis N, Gray A. The economic burden of back pain in the UK. *Pain.* Jan 2000;84(1):95-103.

37. Chan AW, Tetzlaff JM, Altman DG, et al. SPIRIT 2013 statement: defining standard protocol items for clinical trials. *Annals of internal medicine.* Feb 5 2013;158(3):200-207.

Figure legend list

Figure 1. Flow diagram.

Figure 2. Trend for improvement in reporting of sample size calculation over time.

Figure 3. Completeness of sample size calculation description**.**

Figure 4. Discrepancy between the sample size planned and the sample size randomized.

Table 1. General characteristics of the RCTs**.**

Table 2. Commonly reported elements for sample size calculation.

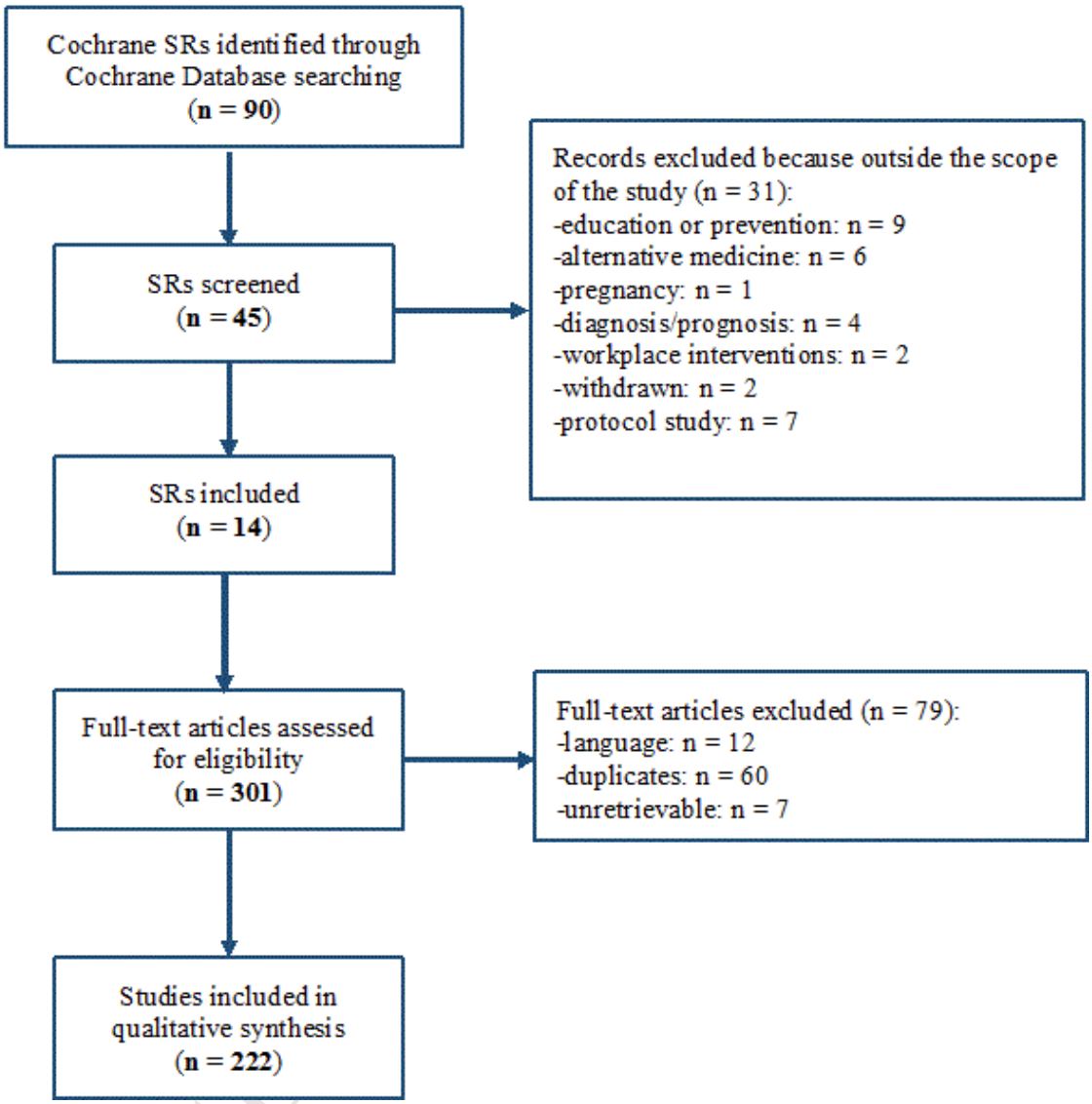
	Frequency (No.)	(%)
No. of countries	32	
USA	42	18.9
UK	29	13.1
The Netherlands	22	9.9
Norway	15	6.8
Sweden	14	6.3
Finland	12	5.4
Australia	10	4.5
Canada	10	4.5
Turkey	10	4.5
No. of journals	78	
Most frequent journals		
Spine	50	22.5
Journal of Manipulative and Physiological Therapeutics	10	4.5
Pain; British Medical Journal; Archives of Physical Medicine and Rehabilitation	9	4.1
Clinical Journal of Pain	8	3.6
No funding reported, no. (%)	97	43.7
	median	**Range**
No. of authors, median (IQR)	5	1-12
Year of publication of trial report, median (IQR)	2000	1968-2013

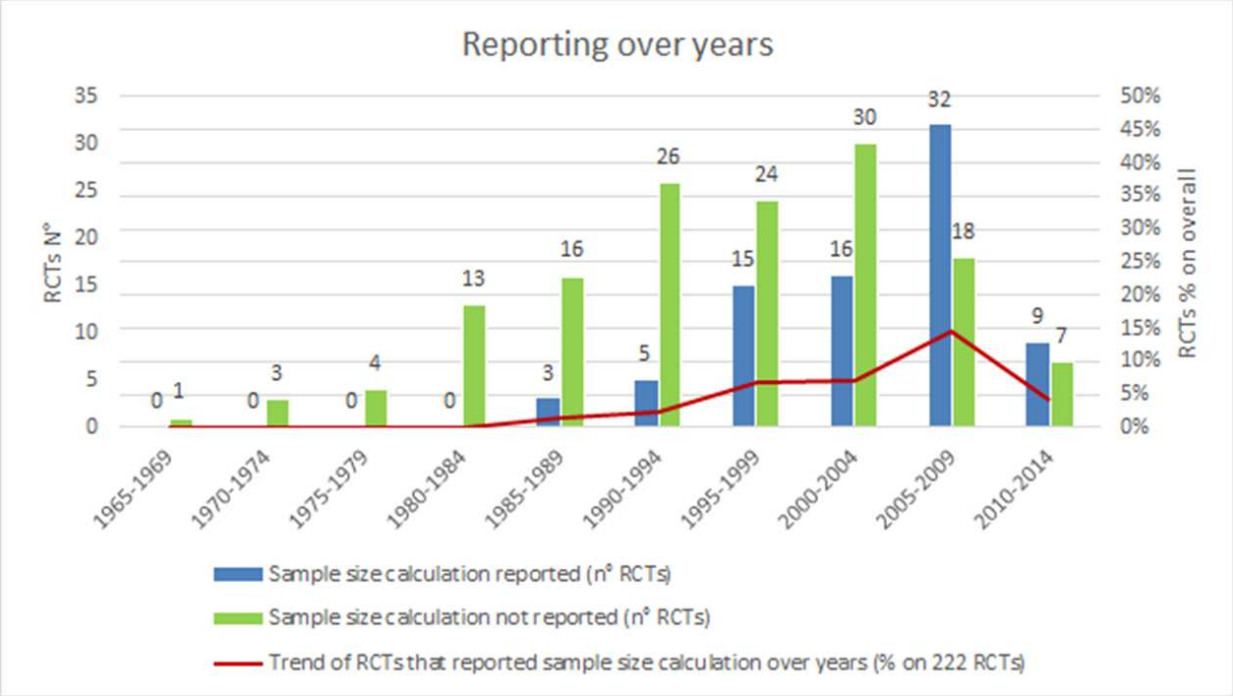
Table 1. General characteristics of the RCTs**.**

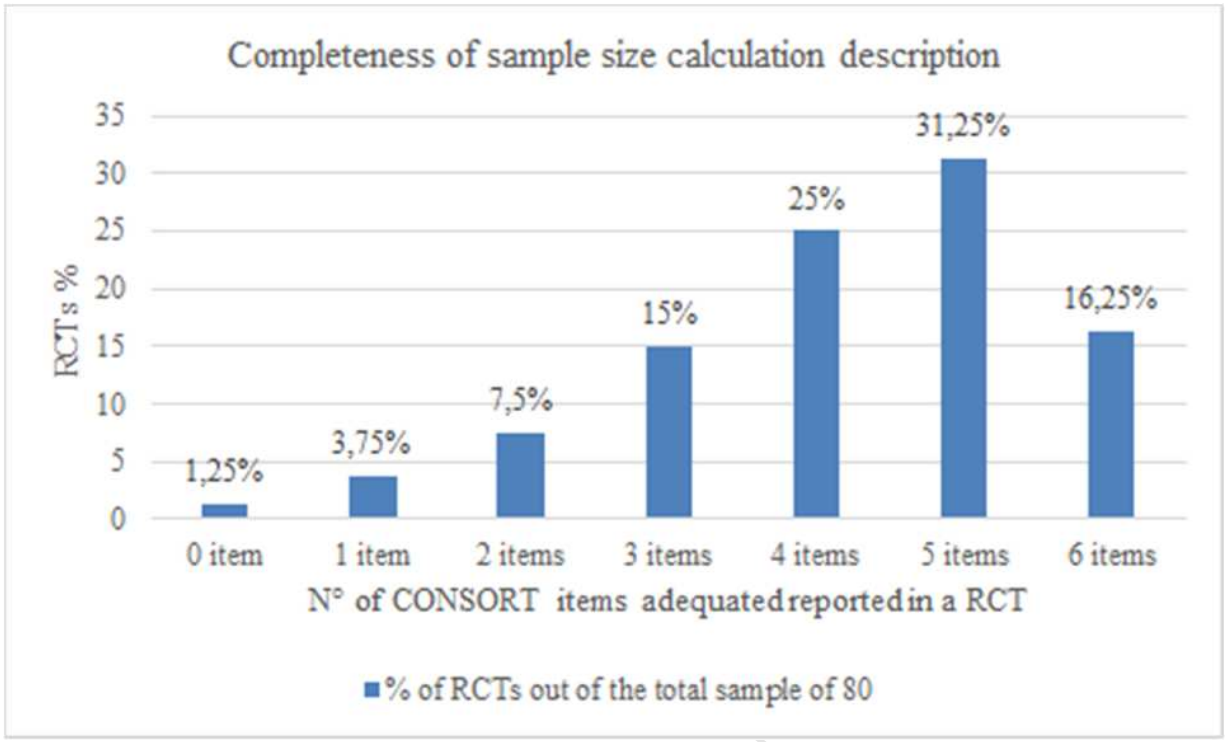
Table 2. Commonly reported elements for sample size calculation.

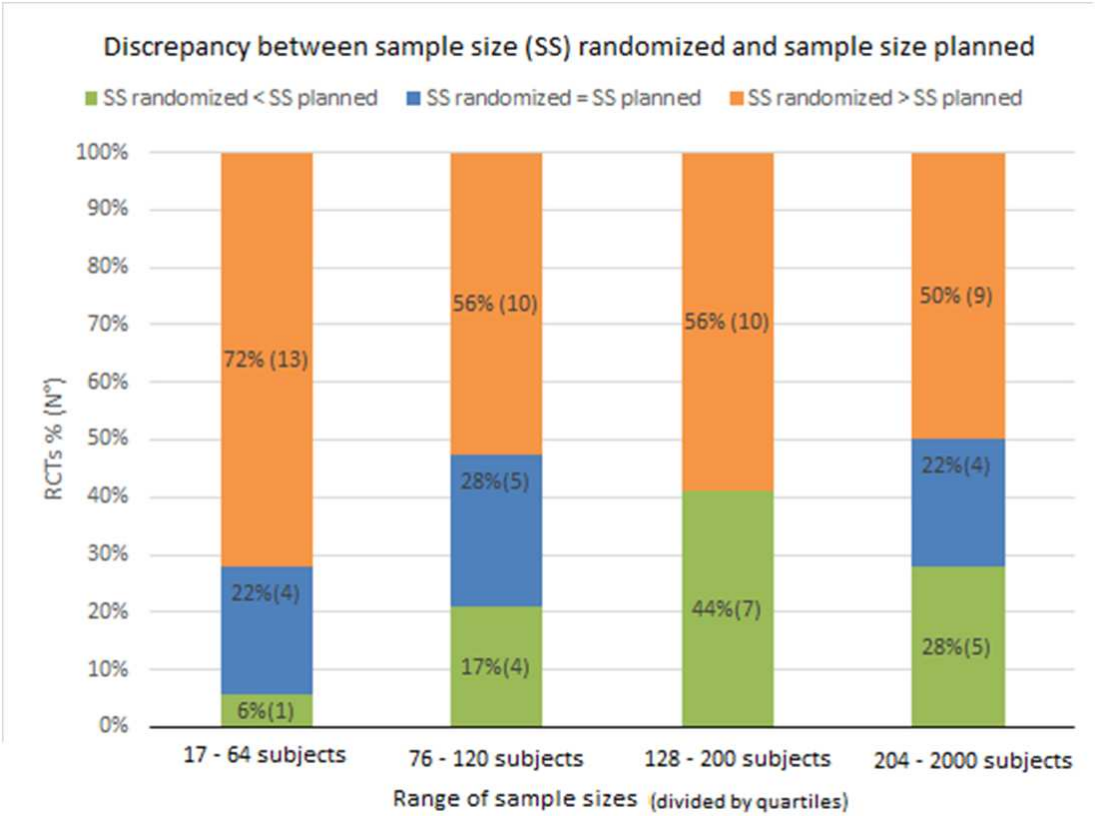
Sample size calculation elements	No. (%)
Level of significance	
Alpha (type I error)	68 (85)
Power	
Beta (type II error)	10 (12.5)
1 - Beta	66 (82.5)
Total	73 (91.3)
Assumption for treatment effect	
MID*	37 (46.3)
Effect Size	9 (11.3)
Other (i.e., reduction in %)	24 (30)
Total	69 (86.3)
Assumption for variability	
Standard deviation	28 (35)
Other (i.e., variance)	7 (8.8)
Total	35 (43.8)
Correction for losses to follow-up	26 (32.5)
Outcome considered for sample calculation	
Disability	34 (42.5)
Pain	18 (22.5)
Other (i.e., recovery rate, work days)	19 (23.8)
Total	63 (78.8)

*MID denotes minimal important difference


Cochrane SRs identified through Cochrane Database searching (n = 90)
SRs screened (n = 45)
Records excluded because outside the scope of the study (n = 31):
-education or prevention: n = 9
-alternative medicine: n = 6
-pregnancy: n = 1
-diagnosis/prognosis: n = 4
-workplace interventions: n = 2
-withdrawn: n = 2
-protocol study: n = 7
SRs included (n = 14)
Full-text articles assessed for eligibility (n = 301)
Full-text articles excluded (n = 79):
-language: n = 12
-duplicates: n = 60
-unretrievable: n = 7
Studies included in qualitative synthesis (n = 222)







Key findings

Numerous RCTs on rehabilitation interventions for mechanical low-back pain, published between the 1960s and the present, failed to report a priori sample size calculation, describing a poor adherence to the CONSORT statement recommendations.

What this adds to what was known

This is the first article that evaluate sample size reporting for each of the CONSORT 2010 recommended descriptive elements in RCTs on low back pain's rehabilitation.

Low-back pain is an increasingly common health problem with a substantial socio-economic burden: despite the call for evidence-based interventions, a lack of methodological quality in rehabilitation RCTs exists.

What is the implication, what should change now

To ensure high quality in conducting a clinical trial, researchers should be mostly encouraged to use international guidelines whereas journal editors and peer reviewers should impose stricter criteria for adequate and transparent reporting.