

Completeness of Outcomes Description Reported in Low Back Pain Rehabilitation Interventions: A Survey of 185 Randomized Trials

Silvia Gianola, MSc, PT;^{*†} Pamela Frigerio, MSc, PT;[‡] Michela Agostini, MSc, PT;[§]
 Rosa Bolotta, MSc, PT;[¶] Greta Castellini, MSc, PT;^{†**} Davide Corbetta, PT;^{††}
 Monica Gasparini, MD, MSc;^{‡‡} Paolo Gozzer, PT;^{§§} Erica Guariento, PT;^{¶¶}
 Linda C. Li, MSc, PhD, PT;^{***†††} Valentina Pecoraro, BSc, MSc;[‡] Valeria Sirtori, PT;^{††}
 Andrea Turolla, MSc, PT;[§] Anita Andreano, MD, MSc;^{*} Lorenzo Moja, MD, MSc, PhD^{†**}

ABSTRACT

Purpose: To assess reporting completeness of the most frequent outcome measures used in randomized controlled trials (RCTs) of rehabilitation interventions for mechanical low back pain. **Methods:** We performed a cross-sectional study of RCTs included in all Cochrane systematic reviews (SRs) published up to May 2013. Two authors independently evaluated the type and frequency of each outcome measure reported, the methods used to measure outcomes, the completeness of outcome reporting using a eight-item checklist, and the proportion of outcomes fully replicable by an independent assessor. **Results:** Our literature search identified 11 SRs, including 185 RCTs. Thirty-six different outcomes were investigated across all RCTs. The 2 most commonly reported outcomes were pain ($n = 165$ RCTs; 89.2%) and disability ($n = 118$ RCTs; 63.8%), which were assessed by 66 and 44 measurement tools, respectively. Pain and disability outcomes were found replicable in only 10.3% ($n = 17$) and 10.2% ($n = 12$) of the RCTs, respectively. Only 40 RCTs (21.6%) distinguished between primary and secondary outcomes. **Conclusions:** A large number of outcome measures and a myriad of measurement instruments were used across all RCTs. The reporting was largely incomplete, suggesting an opportunity for a standardized approach to reporting in rehabilitation science.

Key Words: data reporting; low back pain; outcome measures; randomized controlled trials, as topic; rehabilitation; survey.

RÉSUMÉ

Objectif : évaluer l'exhaustivité des déclarations en ce qui concerne les mesures de résultats utilisées le plus fréquemment dans les essais cliniques randomisés (ECR) portant sur les interventions en réadaptation pour les douleurs chroniques au bas du dos. **Méthodes :** nous avons mené une étude transversale des ECR inclus dans toutes les revues systématiques Cochrane publiées jusqu'en mai 2013. Deux auteurs ont indépendamment évalué : la nature et la fréquence de chacune des mesures de résultats rapportées, les méthodes utilisées pour effectuer ces mesures, l'exhaustivité des déclarations de résultats (à l'aide d'une liste de contrôle en 8 points) et la proportion des résultats qui peuvent être complètement reproduits par un évaluateur indépendant. **Résultats :** notre recension de la littérature a identifié 11 revues systématiques comprenant un total de 185 ECR. Trente-six résultats différents ont été étudiés dans l'ensemble des essais cliniques. Les deux résultats les plus fréquemment rapportés étaient la douleur ($n = 164$ ECR; 89,2%) et l'incapacité ($n = 118$; 63,8%), qui ont été évalués respectivement par 66 et 44 instruments de mesure. Les résultats relatifs à la douleur et à l'incapacité se sont avérés reproductibles dans seulement 10,3% ($n = 17$) et 10,2% ($n = 12$), respectivement, des essais cliniques. Seuls 40 (21,6%) des ECR ont fait la distinction entre le résultat principal et les résultats secondaires. **Conclusion :** un grand nombre de mesures de résultats et d'instruments de mesure ont été utilisés dans l'ensemble des ECR. Les déclarations sont pour la plupart incomplètes; il pourrait y avoir là une occasion de mettre au point une approche standardisée pour la communication des résultats en science de la réadaptation.

From the: ^{*}Center of Biostatistics for Clinical Epidemiology, School of Medicine and Surgery, University of Milano-Bicocca, Monza; [†]Clinical Epidemiology Unit, IRCCS Orthopaedic Institute Galeazzi; [‡]Child and Adolescent Neuropsychiatric Unit, ASST Grande Ospedale Metropolitano Niguarda; [¶]Physiotherapy Service, National Institute of Injury Insurance; ^{**}Department of Biomedical Sciences for Health, University of Milan; ^{††}Functional Recovery Unit, Ospedale San Raffaele, Milan; [§]Kinematics and Robotics Laboratory, IRCCS Fondazione Ospedale San Camillo, Venice; ^{‡‡}Department of Rehabilitation, Azienda sanitaria locale (Local Health Unit) Biella; ^{§§}APSS Tn, Villa Igea, Trento; ^{¶¶}La Quiete Casa di Cura, Varese, Italy; ^{***}Department of Physical Therapy, University of British Columbia; ^{†††}Arthritis Research Centre of Canada, Vancouver.

Correspondence to: Silvia Gianola, Clinical Epidemiology Unit, IRCCS Orthopaedic Institute Galeazzi, Milan, Italy; silvia.gianola@grupposandonato.it.

Contributors: All authors designed the study; collected, analyzed, and interpreted the data; drafted or critically revised the article; and approved the final draft.

Competing Interests: None declared.

This work was supported by the Italian Ministry of Health.

Physiotherapy Canada 2016; 68(3);267–274; doi:10.3138/ptc.2015-30

Randomized controlled trials (RCTs) evaluate the effectiveness of an intervention, which depends on the population included, the characteristics of the intervention, the comparison performed, and the chosen outcome measure. All of these elements need to be carefully evaluated when planning and interpreting research projects.

Researchers make decisions about what outcome to measure in a trial. The type of outcome measure influences both the magnitude of the clinically important difference attributable to an intervention and the definition of a successful outcome.¹ In fact, success depends on demonstrating the statistically significant difference and the clinical relevance of the benefit. Also, the chosen outcome measure will influence the sample size required for a trial² and the length of follow-up needed to accumulate a sufficient number of events from which to draw a firm conclusion. Finally, the choice of the reported outcome and its measure are subject to selective outcome reporting bias (i.e., the outcome that results in statistical significance is the one reported in a publication).³ All of these considerations apply to the rehabilitation field, in which the need to evaluate diverse research objectives and different dimensions leads to the use of various outcomes and outcome measures.⁴

Even when RCTs use similar populations, differences in outcome measures make it difficult to compare study results and assess the relative magnitude of treatment effects among various studies.⁵ For instance in a study from Mohseni-Bandpei and colleagues,⁶ when spinal manipulative therapy was compared with any other intervention for chronic low back pain (LBP), the functional 100-point Oswestry Disability Index was associated with a large and statistically significant absolute improvement. However, when the same outcome was evaluated by Bronfort and colleagues⁷ using the 24-point Roland Morris Disability Questionnaire, they reported an absolute smaller difference that was not statistically significant.

Furthermore, many outcome measures in rehabilitation are multidimensional, producing several domain-specific scales in patient-reported outcomes (PROs), including health-related quality of life (HRQOL) and symptoms such as pain or fatigue.^{8,9} Unfortunately, evidence has shown that in some trials, the quality of PRO data can be undermined by inconsistencies in data collection¹⁰ and, in particular, by high rates of missing data;¹¹ this adversely affects the integrity and usefulness of such data in clinical practice.¹²

The adequacy of the assessment for the chosen outcome can be evaluated only from its reporting. Recently, the development of the Consolidated Standards of Reporting Trials (CONSORT) PRO extension⁹ has introduced important recommendations about PROs: a precise definition of the outcome, including whether it was a primary or secondary outcome, and how it was measured, specifying the setting and the timing of the assessment.^{9,13} Also, the Standard Protocol Items: Recommen-

ations for Interventional Trials (SPIRIT) initiative¹⁴ has promoted trial conduct and reporting in its protocols.¹² Completeness of outcome reporting is important because it allows clinicians and other health professionals to apply the treatment in practice, guides researchers using previous research to shape future studies, and informs those analyzing the clinical trial data in meta-analyses to draw more reliable conclusions.

The aim of our study was to evaluate the completeness of reporting of outcomes that are most commonly used in RCTs examining interventions for LBP. To pursue this goal, we first determined the type and frequency of outcomes used. We then determined and examined the completeness of the reporting of the four most commonly used outcomes to determine whether there was a relationship with the year in which the trial was published. We hypothesized that outcomes would be reported more thoroughly in recently published trials supported by various initiatives promoting reporting, such as CONSORT PRO and SPIRIT.^{9,14,15} Finally, we examined how complete the description was of the blinding of the outcome assessment.

METHODS

Registered protocol

We registered the present study in the Core Outcome Measures in Effectiveness Trials (COMET) database,¹⁶ in agreement with the COMET initiative.¹⁷

Eligibility criteria and study selection

We searched all systematic reviews (SRs) published up to May 2013 in the Cochrane Database of Systematic Reviews, using the terms “back pain” and “rehabilitation” in adult treatments. Interventions other than therapeutic rehabilitation (e.g., education) and those based on a subgroup population (e.g., spondylolisthesis) were excluded.

From the eligible SRs, we extracted all RCTs published in English, Italian, Spanish, or French. Three authors (SG, PF, GC) independently screened the SRs (title and abstract) for eligibility and subsequently reviewed all identified RCTs. Disagreements were resolved by negotiation among the authors.

Data collection and definitions

We designed an outcome extraction form, then refined it after conducting the first 60 trials, based on the problems identified. Using DistillerSR (Evidence Partners, Ottawa, ON), a web-based, password-protected database for data extraction, six pairs of independent researchers trained in SR methodology extracted study characteristics—such as information concerning the study population, intervention, control, sample size, number of reported outcomes and their assessment, and funding—from the included RCT full text. (See Appendix 1 online.) We further recorded whether each RCT distinguished between primary and secondary outcomes.

Table 1 Characteristics of Pain, Disability, Range of Motion, and HRQOL Outcomes

	Most reported outcome, no. (%)			
	Pain	Disability	Range of motion	HRQOL
Studies (<i>n</i> = 185)	165 (89.2)	118 (63.8)	72 (39.9)	45 (24.3)
Studies citing this as the primary outcome (<i>n</i> = 31)*	13 (41.9)	19 (61.3)	0 (0)	1 (3.2)
Instruments				
Self-reported	62 (88.6)	43 (100)	0 (0)	19 (100)
Total reported	70	43	41	19

*Among 40 studies that identified the primary outcome, we considered only the 31 trials reporting one outcome as the primary outcome in the Methods section or using it for sample size calculation. This included 9 trials that used combined outcomes and 7 trials in which the primary outcome was not one of the four most reported outcomes.

HRQOL = health-related quality of life.

We defined the primary outcome as being adequately reported when only one outcome, even if composite, had been indicated as primary in the Methods section or had been used in calculating the sample size. If we were unsure about the primary outcome (e.g., more than one outcome defined as primary, no indication that the outcome was used to calculate sample size in the presence of multiple outcomes), it was considered to be not adequately reported.

After determining the four most frequently reported outcomes, we assessed the completeness of reporting using an eight-item checklist that we had developed specifically for this project; the items are defined in Table 1. They were selected from established opinion on what aspects of the methodology should be reported.^{1,9,18–21} The Methods and Results sections of each trial were reviewed, and we decided whether each of the eight items was reported or not reported. An outcome was considered to be fully reported if all of the items were present. We also analyzed changes in the completeness of reporting for each outcome over time. Finally, because blinding is one of the most important procedures to protect against bias in an RCT,²² we investigated its reporting by determining the frequency of blinding across all included RCTs. A trial was considered to be blinded, unblinded, or unclear on the basis of the information provided in the article.²² When blinding was reported, we specified the level: participants, trial investigators, outcome assessors, or data analysts.

Statistical analysis

Completeness of reporting for the four most frequent outcomes was described, for every item on the checklist, by the proportion of RCTs adequately reporting the item. For every outcome, univariate logistic regression models were used to investigate the impact on each item (dependent binary variable) of publication year (continuous independent variable). We modelled the proper functional form of year using polynomial terms. For items with a significant quadratic term—representing a decreasing and then increasing proportion of adequately reported RCTs with publication year—we estimated the

linear effect of publication year for the most recent time period. To do so, we fitted a new model, including just the linear term, only on the studies published after the curvature point. The results of the logistic regressions are presented graphically and as 10-year odds ratios (ORs)—that is, the relative increase or decrease in the probability that a study will report the item for any 10-year increment in publication year—and their corresponding 95% CIs. All tests were performed two-sided, with a significance level of 0.05. All analyses were performed with R (R Foundation for Statistical Computing, Vienna, Austria).

RESULTS

Selection of studies

Eleven Cochrane SRs met our criteria, and they identified a total of 220 RCTs. We removed any trials that were duplicates; in a language other than French, English, Italian, or Spanish; or unavailable; this left 185 RCTs for analysis. A more thorough description of the study selection is presented in Figure S1 (online Appendix 2).

How many outcomes and measurements are reported in the published randomized controlled trials?

Overall, 36 outcomes were reported more than once across the studies, and more than 100 outcomes were reported only once. The outcomes most frequently reported were pain (89.2%), disability (63.8%), range of motion (38.9%), and HRQOL (24.3%). (See Table 2.) We found 70 instruments used to assess pain (e.g., visual analogue scale [VAS], numeric rating scale), 43 to assess disability (e.g., Roland Morris Disability Questionnaire, Oswestry Disability Index), 41 to assess range of motion (e.g., Modified Schober), and 19 to assess HRQOL (e.g., SF-36 health survey, Euro-Qol). Disability and HRQOL were assessed using self-report measures, and range of motion was always assessed with a clinical assessment. Pain was investigated using either clinical measures (e.g., pressure pain was measured using a commercial algometer applied by the health professional) or self-reported measures (e.g., a patient-reported scale such as a VAS).

Table 2 Reporting of Outcome Assessment

Item	Description	No. (%) of trials reporting the item for each outcome			
		Pain (<i>n</i> = 165)	Disability (<i>n</i> = 118)	Range of motion (<i>n</i> = 72)	HRQOL (<i>n</i> = 45)
Bias	Was outcome assessment blinded?	76 (46.1)	53 (44.9)	41 (56.9)	17 (37.8)
Data collection	Were data collection methods clearly specified?	49 (29.7)	41 (34.7)	17 (23.6)	18 (40.0)
Assessor	Was the assessor of the outcome stated?	81 (49.1)	68 (57.6)	43 (59.7)	26 (57.8)
Timelines	Was the follow-up schedule detailed?	145 (87.9)	103 (87.3)	63 (87.5)	39 (86.7)
Reliability	Were the validity and reliability of the instrument provided (in the study or in reference to a validation study)?	112 (67.9)	104 (88.1)	43 (59.7)	41 (91.1)
Properties	Was the process of measurement of the outcome fully described?	124 (75.2)	77 (65.2)	48 (66.7)	24 (53.3)
Instrument	Was the specific instrument used to measure the outcome reported?	150 (90.9)	112 (94.9)	58 (80.6)	44 (97.8)
Concept	Was the outcome clearly defined?	129 (78.2)	90 (76.2)	61 (84.7)	31 (68.9)

HRQOL = health-related quality of life.

Table 3 Odds Ratio for the Four Most Reported Outcomes Being Reported versus Not Reported for Any 10-Year Increment in Publication Year

Item	10-year OR (95% CI)			
	Pain	Disability	Range of motion	HRQOL
Bias	1.78 (1.17, 2.72)	2.54 (1.40, 4.59)	1.24 (0.77, 1.99)	2.41 (0.84, 6.91)
Data collection	1.78 (1.17, 2.70)	1.87 (1.04, 3.37)	1.25 (0.71, 2.22)	2.72 (0.94, 7.87)
Assessor	1.37 (0.97, 1.93)	1.45 (0.87, 2.40)	1.19 (0.74, 1.92)	2.39 (0.89, 6.39)
Timelines	1.34 (0.82, 2.19)	1.28 (0.63, 2.60)	1.14 (0.57, 2.28)	0.59 (0.13, 2.59)
Reliability	2.33 (1.57, 3.46)	5.64 (2.38, 13.34)	1.42 (0.87, 2.30)	0.95 (0.18, 4.86)
Properties	2.02 (1.35, 3.02)	1.07 (0.64, 1.81)	1.78 (1.06, 3.00)	0.66 (0.26, 1.70)
Instrument	2.36 (1.33, 4.17)	4.90 (1.59, 15.09)	4.41 (1.65, 11.78)	—*
Concept	1.48 (0.99, 2.20)	1.47 (0.84, 2.59)	1.70 (0.90, 3.22)	1.37 (0.51, 3.69)

*It was not possible to fit the logistic model for Instrument because the proportion was almost always close to 1.

HRQOL = health-related quality of life.

Did the authors specify primary and secondary outcomes?

Of the 185 RCTs, 40 (21.6%) distinguished between primary and secondary outcomes in the Methods section, and 31 (77.5%) of these 40 trials adequately identified the primary outcome. Table 2 provides additional information on the characteristics of each outcome. The adequate reporting of the primary outcome appeared to have improved over time: from 0% before 1994 (when no study reported the primary outcome) to 8.6% (3 of 35) between 1995 and 1999, 26.8% (11 of 41) between 2000 and 2004, and 44.75% (17 of 38) between 2005 and 2010.

Completeness of outcome reporting

We evaluated the completeness of reporting for the four most frequently reported outcomes: pain, disability, range of motion, and HRQOL. Table 1 presents the proportion of RCTs reporting each item.

The items with the most thorough reporting were the type of instrument, timeline and follow-up schedule, and reliability of the instrument. The items less frequently judged as adequately reported were the methods used

for data collection and the methods used during the process to protect against bias. This was consistent across all four outcomes.

For the four most frequently reported outcomes, only a few trials reported all items: 10.3% for pain (17 of 165), 10.2% for disability (12 of 118), 5.5% for range of motion (4 of 72), and 6.7% for HRQOL (3 of 45). The majority of the RCTs provided insufficient detail on these four outcomes to allow them to be fully understood and replicated in future trials.

In the RCTs that included all 4 (*n* = 5) or 3 (*n* = 59) of the most commonly reported outcomes, none was successful in adequately reporting all outcomes (eight of eight items). In RCTs featuring 2 of the most frequently reported outcomes (*n* = 86), only 6% of trials successfully reported both.

The four outcomes first appeared in the literature at different times: pain and range of motion in 1968, disability in 1977, and HRQOL only in 1988. For each outcome, Table 3 reports the OR of each item being reported versus not reported for any 10-year increment

Table 4 Reporting of the Use and Level of Blinding for the Four Most Reported Outcomes

	No. (%) of RCTs			
	Pain	Disability	Range of motion	HRQOL
Blinding*	165 (89.2)	118 (63.8)	72 (38.9)	45 (24.3)
Unclear	69 (41.8)	52 (44.1)	24 (33.3)	16 (35.6)
None	14 (8.5)	13 (11.0)	5 (6.9)	7 (15.6)
Yes	82 (49.7)	53 (44.9)	43 (59.7)	22 (48.9)
Type of blinding†				
Participants	24 (29.3)	15 (28.3)	11 (25.6)	6 (27.3)
Trial investigators	20 (24.4)	14 (26.4)	11 (25.6)	4 (18.2)
Outcome assessors	69 (84.1)	40 (75.5)	37 (86.0)	16 (72.7)
Data analysts	9 (11.0)	7 (13.2)	4 (9.3)	4 (18.2)

Note: Percentages may total less than or more than 100 because of rounding.

*Percentages are of the total number of RCTs ($n = 185$).

†Percentages are of the total number of RCTs reporting a blinded assessment (Yes row). A trial could have adopted one or more types of blinding (e.g., one in which both trial investigators and assessors were blinded).

RCT = randomized controlled trial; HRQOL = health-related quality of life.

in publication year. In Figures S2–S5 (online Appendix), data points at 5-year intervals (except for the first time interval, which varied according to outcome) represent the proportion of RCTs (y -axis) that reported each of the eight items; eight continuous curves represent the relationship between the reporting frequency of each item and the year of publication, as estimated from the logistic model and back-transformed on the proportion scale.

The outcomes showing a statistically significant improvement in reporting over time were as follows: for pain, instrument (10 y OR = 2.4, $p = 0.003$), properties (10 y OR = 2.0, $p = 0.001$), reliability (10 y OR = 2.3, $p < 0.001$), data collection (10 y OR = 1.8, $p = 0.007$), and bias (10 y OR = 1.8, $p = 0.007$), with the latter having a significant ascending trend only from 1980; for disability, instrument (10 y OR = 4.9, $p = 0.006$), reliability (10 y OR = 5.6, $p < 0.001$), data collection (10 y OR = 1.9, $p = 0.037$), and bias (10 y OR = 2.5, $p = 0.002$); and for range of motion, properties (10 y OR = 1.8, $p = 0.030$). Instrument had a significant ascending trend only from 1980 (10 y OR = 4.4, $p = 0.03$), and for HRQOL, none of the items had a significant change over time. None of the outcomes had a statistically significant decrease in reporting over time.

Table 4 details the reporting of the use and level of blinding for the four most frequent outcomes. The use of blinded assessment was adequately reported in about half of the RCTs and unclearly reported in fewer than half, ranging from 33.3% (for range of motion) to 44.1% (for disability) of the trials. The percentage of trials explicitly reporting no blinding varied between 6.9% (for range of motion) and 15.6% (for HRQOL).

For the four most commonly reported outcomes, blinding was more frequently performed for the outcome assessors (the persons who determined the outcome

measurement—e.g., the participants, researchers, or independent assessors), ranging from 72.7% (for HRQOL) to 86.0% (for range of motion). This was followed by participants, ranging from 25.6% (for range of motion) to 29.3% (for pain); trial investigators, ranging from 18.2% (for HRQOL) to 26.4% (for disability); and data analysts, ranging from 9.3% (for range of motion) to 18.2% (for HRQOL).

DISCUSSION

Outcome assessment is not adequately reported in RCTs for LBP interventions. First, we identified numerous outcomes and outcome measurements used to evaluate rehabilitation of mechanical LBP. Overall, 36 outcomes were reported more than once across the studies, and more than 100 outcomes were reported only once. Second, only one-fifth of the trials declared a primary outcome. Finally, only about 60% of the trials declared the methods used to protect outcome assessment against bias (i.e., blinded assessment).

It has been claimed that insufficient attention is paid to outcome measurement in clinical trials.²³ The large heterogeneity in outcome measurement in LBP is not new.⁴ It can increase the gap among scientists, clinicians, and patients because they are reasonably skeptical about accepting a research field that proposes dozen of outcomes to be clinically relevant. Variations in the measurement of the same outcome can often explain apparent discrepancies in results across similar studies.¹ However, the poor reporting of outcomes can lead to such important differences being overlooked. Heterogeneity in outcome measures also complicates meta-analyses.²⁴ When different instruments are used, standardized mean difference (SMD) is usually adopted as the summary statistic to reflect the effect size.²⁵ SMD is defined as the ratio of

mean to SD of the difference of two random values, respectively, from two groups. However, it expresses the intervention effect in SD units rather than in the original units of measurement, with the value of an SMD depending on both the size of the effect and the SD of the outcomes. This approach has two main limitations: Health professionals do not have an intuitive sense of the importance of the effect if it is expressed as an SD unit, and the same effect will assume different SMD values if population heterogeneity differs across eligible trials.^{26,27}

Difficulties caused by heterogeneity in outcome measurement could be addressed by the development and use of a core outcome set (COS). This is a scientifically agreed-on set of outcomes, and it has to be reported as a minimum in RCTs conducted in a specific area of clinical practice.²⁸ The most successful example of COS is arthritis trials using Outcome Measures in Rheumatology (OMERACT) initiatives.^{20,29} Other examples of its use in fields close to rehabilitation are chronic post-surgical pain after knee replacement³⁰ and hip fracture.³¹

In our study, we found that pain, disability, range of motion, and HRQOL are the most widely used outcomes. The identification and recognition of the most frequently used outcomes across clinical trials represent a first step in the development of an LBP COS. The second step is to evaluate the relevance of these outcomes for patients, health care practitioners, regulators, industry representatives, and policymakers; joining diverse stakeholders in the endeavour to reach a consensus is increasingly well accepted as the future of collaborative research.³²

In 1998, after an expert panel discussion held at the second international LBP Forum (in The Hague, Netherlands), a proposal was published for a standardized six-item set of outcomes in LBP clinical research: pain symptoms, function, well-being, disability, disability (social role), and satisfaction with care.⁵ More recently, a group of researchers has been updating these recommended domains for LBP clinical research,³³ following the methodological guidance of COMET²³ and OMERACT.²⁹

The small proportion of trials that distinguish between primary and secondary outcomes presents another major issue. Not indicating a single primary outcome can lead to outcome reporting bias. Recommendations for intervention trial protocols were published in 2013 by the SPIRIT initiative. The SPIRIT checklist considers a full description of the study planning, including the identification of the primary and secondary outcomes.¹⁴ A greater adherence to these recommendations should be pursued.

Although overall reporting has improved over time, items are still insufficiently reported. The incompleteness of reporting of other outcome dimensions leads to multiple biases, such as performance or detection bias.²²

The scarcity of blinding is also worrying; blinding is more difficult to achieve in rehabilitation interventions

than in pharmacological trials because patients and health care providers are aware of the allocated treatment.³⁴ Although these subjects can rarely be blinded, it is usually possible to blind the outcome assessors to ensure unbiased ascertainment of outcomes, especially in the presence of subjective outcomes.³⁵ We invite authors of the rehabilitation literature, journal editors, and reviewers to improve their adherence to CONSORT PRO, promoting blinded assessment to reduce uncertainty about the methods used to assess outcomes.

This study has some limitations that should be kept in mind when interpreting the results. For example, to determine whether the overall reporting was satisfactory, we selected the highest possible threshold for adequate reporting (i.e., all items on the checklists). Lower thresholds would have increased the number of compliant records, but we judged that for a study to be truly replicated, all items needed to be present. Completeness of reporting may have been influenced not only by the dimension of the outcome (e.g., pain) or the measure used (e.g., Oswestry Disability Index) but also by other merits and limitations (e.g., binary vs. continuous, interpretability, relevance, statistical significance). Moreover, we did not explore the implications of poor reporting when the results of the RCTs were disseminated to patients.

To capture the selective outcome reporting bias, we would need to study discrepancies between the registered protocol and its corresponding full text. Because our sample dates to 1970, it is difficult to detect potential bias because the widespread registration of protocols only began in the past 10 years. Finally, when an outcome was assessed using a multidimensional scale (e.g., the Oswestry Disability Index, which encompasses pain and disability), we arbitrarily retained only the most inclusive dimension construct (e.g., disability).

CONCLUSION

A large number of outcomes and a multitude of instruments have been used in RCTs examining physical interventions for LBP in adults. The thoroughness with which the outcome measures are described has improved over time but remains incomplete. Our findings suggest that ongoing attention to the description of outcome measures is needed from authors, peer reviewers, and journal editors. A COS for LBP may help initiatives such as SPIRIT and CONSORT PRO.

KEY MESSAGES

What is already known on this topic

There is heterogeneity in outcomes and outcome measures across low back pain (LBP) trials; this may affect consistency of reporting and completeness of the description and make it difficult to perform systematic reviews.

What this study adds

We quantified the heterogeneity in outcome and outcome measures reported in LBP rehabilitation trials. We also found that the reporting of the outcome measures, assessed by eight items, has overall improved over time. However, some aspects of the reporting are still incomplete. We call for the definition of a Core Outcome Set with a complete description of the outcome assessment in this field.

REFERENCES

- Coster WJ. Making the best match: selecting outcome measures for clinical trials and outcome studies. *Am J Occup Ther.* 2013;67(2):162–70. <http://dx.doi.org/10.5014/ajot.2013.006015>. Medline:23433270
- Leon AC, Marzuk PM, Portera L. More reliable outcome measures can reduce sample size requirements. *Arch Gen Psychiatry.* 1995;52(10):867–71. <http://dx.doi.org/10.1001/archpsyc.1995.03950220077014>. Medline:7575107
- Chan AW, Hróbjartsson A, Haahr MT, et al. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA.* 2004;291(20):2457–65. <http://dx.doi.org/10.1001/jama.291.20.2457>. Medline:15161896
- Chapman JR, Norvell DC, Hermsmeyer JT, et al. Evaluating common outcomes for measuring treatment success for chronic low back pain. *Spine.* 2011;36(21 Suppl):S54–68. <http://dx.doi.org/10.1097/BRS.0b013e31822ef74d>. Medline:21952190
- Deyo RA, Battie M, Beurskens AJ, et al. Outcome measures for low back pain research. A proposal for standardized use. *Spine.* 1998;23(18):2003–13. <http://dx.doi.org/10.1097/00007632-199809150-00018>. Medline:9779535
- Mohseni-Bandpei MA, Critchley J, Staunton T, et al. A prospective randomised controlled trial of spinal manipulation and ultrasound in the treatment of chronic low back pain. *Physiotherapy.* 2006;92(1):34–42. <http://dx.doi.org/10.1016/j.physio.2005.05.005>.
- Bronfort G, Goldsmith CH, Nelson CF, et al. Trunk exercise combined with spinal manipulative or NSAID therapy for chronic low back pain: a randomized, observer-blinded clinical trial. *J Manipulative Physiol Ther.* 1996;19(9):570–82. Medline:8976475
- Food and Drug Administration. Guidance for industry—patient-reported outcome measures: use in medical product development to support labeling claims. Silver Spring (MD): Office of Communications, Division of Drug Information, Center for Drug Evaluation and Research, Food and Drug Administration; 2009.
- Calvert M, Blazeby J, Altman DG, et al.; CONSORT PRO Group. Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. *JAMA.* 2013;309(8):814–22. <http://dx.doi.org/10.1001/jama.2013.879>. Medline:23443445
- Kyte D, Ives J, Draper H, et al. Inconsistencies in quality of life data collection in clinical trials: a potential source of bias? Interviews with research nurses and trialists. *PLoS One.* 2013;8(10):e76625. <http://dx.doi.org/10.1371/journal.pone.0076625>. Medline:24124580
- Fielding S, MacLennan G, Cook JA, et al. A review of RCTs in four medical journals to assess the use of imputation to overcome missing data in quality of life outcomes. *Trials.* 2008;9(1):51. <http://dx.doi.org/10.1186/1745-6215-9-51>. Medline:18694492
- Calvert M, Kyte D, Duffy H, et al. Patient-reported outcome (PRO) assessment in clinical trials: a systematic review of guidance for trial protocol writers. *PLoS One.* 2014;9(10):e110216. <http://dx.doi.org/10.1371/journal.pone.0110216>. Medline:25333995
- McDowell I, Newell C. *Measuring health. A guide to rating scales and questionnaires.* New York: Oxford University Press; 1996.
- Chan AW, Tetzlaff JM, Altman DG, et al. SPIRIT 2013 statement: defining standard protocol items for clinical trials. *Ann Intern Med.* 2013;158(3):200–7. <http://dx.doi.org/10.7326/0003-4819-158-3-201302050-00583>. Medline:23295957
- Boutron I, Moher D, Altman DG, et al.; CONSORT Group. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. *Ann Intern Med.* 2008;148(4):295–309. <http://dx.doi.org/10.7326/0003-4819-148-4-200802190-00008>. Medline:18283207
- COMET Initiative [Internet]. Liverpool: COMET Initiative; 2011–2016 [cited 2015 Jan]. Available from: <http://www.comet-initiative.org/studies/searchresults/>.
- COMET Initiative [Internet]. Liverpool: COMET Initiative; 2011–2016 [cited 2015 Feb 16]. Available from: <http://www.comet-initiative.org/>.
- McDowell I. *Measuring health: a guide to rating scales and questionnaires.* 3rd ed. Oxford: Oxford University Press; 2006. <http://dx.doi.org/10.1093/acprof:oso/9780195165678.001.0001>.
- Schulz KF, Grimes DA. Blinding in randomised trials: hiding who got what. *Lancet.* 2002;359(9307):696–700. [http://dx.doi.org/10.1016/S0140-6736\(02\)07816-9](http://dx.doi.org/10.1016/S0140-6736(02)07816-9). Medline:11879884
- Tugwell P, Boers M; OMERACT Committee. Developing consensus on preliminary core efficacy endpoints for rheumatoid arthritis clinical trials. *J Rheumatol.* 1993;20(3):555–6. Medline:8478872
- Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol.* 2010;63(7):737–45. <http://dx.doi.org/10.1016/j.jclinepi.2010.02.006>. Medline:20494804
- Higgins J, Altman D, Sterne Je. Assessing risk of bias in included studies. In: Higgins JPT, Green S, editors. *Cochrane handbook for systematic reviews of interventions* [Internet]. Version 5.1.0. London: Cochrane Collaboration; 2011 [cited 2005 May 5]. Available from: handbook.cochrane.org.
- Williamson PR, Altman DG, Blazeby JM, et al. Developing core outcome sets for clinical trials: issues to consider. *Trials.* 2012;13(1):132. <http://dx.doi.org/10.1186/1745-6215-13-132>. Medline:22867278
- Kirkham JJ, Gargon E, Clarke M, et al. Can a core outcome set improve the quality of systematic reviews? A survey of the coordinating editors of Cochrane Review Groups. *Trials.* 2013;14(1):21. <http://dx.doi.org/10.1186/1745-6215-14-21>. Medline:23339751
- Deeks J, Higgins J, Altman D. Analysing data and undertaking meta-analyses. In: Higgins JPT, Green S, editors. *Cochrane handbook for systematic reviews of interventions* [Internet]. Version 5.1.0. London: Cochrane Collaboration; 2011 [cited 2015 May 5]. Available from: handbook.cochrane.org.
- Johnston BC, Thorlund K, Schünemann HJ, et al. Improving the interpretation of quality of life evidence in meta-analyses: the application of minimal important difference units. *Health Qual Life Outcomes.* 2010;8(1):116. <http://dx.doi.org/10.1186/1477-7525-8-116>. Medline:20937092
- Johnston BC, Patrick DL, Thorlund K, et al. Patient-reported outcomes in meta-analyses—part 2: methods for improving interpretability for decision-makers. *Health Qual Life Outcomes.* 2013;11(1):211. <http://dx.doi.org/10.1186/1477-7525-11-211>. Medline:24359184
- Ketola S, Lehtinen J, Rousi T, et al. No evidence of long-term benefits of arthroscopic acromioplasty in the treatment of shoulder impingement syndrome: five-year results of a randomised controlled trial. *Bone Joint Res.* 2013;2(7):132–9. <http://dx.doi.org/10.1302/2046-3758.27.2000163>. Medline:23836479
- Tugwell P, Boers M, Brooks P, et al. OMERACT: an international initiative to improve outcome measurement in rheumatology. *Trials.* 2007;8(1):38. <http://dx.doi.org/10.1186/1745-6215-8-38>. Medline:18039364
- Wylde V, MacKichan F, Bruce J, et al. Assessment of chronic post-surgical pain after knee replacement: development of a core outcome set. *Eur J Pain.* 2015;19(5):611–20. Medline:25154614

31. Haywood KL, Griffin XL, Achten J, et al. Developing a core outcome set for hip fracture trials. *Bone Joint J*. 2014;96-B(8):1016–23. <http://dx.doi.org/10.1302/0301-620X.96B8.33766>. Medline:25086115
32. Liberati A. Need to realign patient-oriented and commercial and academic research. *Lancet*. 2011;378(9805):1777–8. [http://dx.doi.org/10.1016/S0140-6736\(11\)61772-8](http://dx.doi.org/10.1016/S0140-6736(11)61772-8). Medline:22098852
33. Chiarotto A, Terwee CB, Deyo RA, et al. A core outcome set for clinical trials on non-specific low back pain: study protocol for the development of a core domain set. *Trials*. 2014;15(1):511. <http://dx.doi.org/10.1186/1745-6215-15-511>. Medline:25540987
34. Boutron I, Guttet L, Estellat C, et al. Reporting methods of blinding in randomized trials assessing nonpharmacological treatments. *PLoS Med*. 2007;4(2):e61. <http://dx.doi.org/10.1371/journal.pmed.0040061>. Medline:17311468
35. Karanicolas PJ, Farrokhyar F, Bhandari M. Practical tips for surgical research: blinding: who, what, when, why, how? *Can J Surg*. 2010;53(5):345–8. Medline:20858381