Journal of Physiotherapy 71 (2025) 1-3

Journal of PHYSIOTHERAPY

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

Editorial

Surrogate endpoints in randomised trials of physiotherapy interventions: the SPIRIT and CONSORT extension checklists for better reporting

Christina DCM Faria [a], Janaine Cunha Polese [a], Anthony Muchai Manyara [b,c], Rod S Taylor [c,d], Oriana Ciani [e], Aline Alvim Scianni [a]

[a] Physiotherapy Department, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil; [b] Global Health and Ageing Research Unit, Bristol Medical School, University of Bristol, Bristol, UK; [c] MRC/CSO Social and Public Health Sciences Unit, School of Health and Wellbeing, University of Glasgow, Glasgow, UK; [d] Robertson Centre for Biostatistics, School of Health and Wellbeing, University of Glasgow, Glasgow, UK; [e] Centre for Research on Health and Social Care Management, SDA Bocconi School of Management, Milan, Italy

Standard guidelines have been systematically developed and are commonly used to guide the completeness and quality of protocols (Standard Protocol Items: Recommendations for Interventional Trials [SPIRIT])[1] and final reports (Consolidated Standards of Reporting Trials [CONSORT])[2] of randomised controlled trials (RCTs). These guidelines have been updated and improved based on experts' consensus, and are widely endorsed for randomised trials,[1–3] including those conducted in the physiotherapy field.[4,5] Recently, a framework that extends the definition and interpretation of surrogate endpoints in trials was developed, comprising the SPIRIT-Surrogate extension checklist[6] and the CONSORT-Surrogate extension checklist.[7] These checklists guide researchers, including those in physiotherapy, with the minimum requirements for protocol development and reporting randomised trials where surrogate endpoints are used as primary outcomes.[6,7] This Editorial aims to disseminate these frameworks and extension checklists in the physiotherapy field.

According to the Physiotherapy Evidence Database (PEDro), over 46,000 randomised trials have been conducted and published in various subdisciplines of physiotherapy, including cardiothoracics, continence/pelvic health, ergonomics/occupational health, gerontology, musculoskeletal, neurology, oncology, orthopaedics, paediatrics and sports.[4,8–11] The results of these trials have helped physiotherapists to manage patients with the evidence-based practice approach.[12,13] Carefully designed, conducted and reported randomised trials are key to measuring intervention efficacy, thereby contributing to evidence-based care.[13,14]

In recent decades, a growing interest in healthcare outcomes that matter to patients – such as body weight, muscle injury, pain, global muscle strength, function, satisfaction and health-related quality of life – has been observed.[3,14–18] These outcomes are commonly known as patient-relevant outcomes and have become common in trials as primary outcomes (or primary endpoints)[3,14–18] (ie, the most important outcome, which is considered in the sample size calculation and in drawing the main conclusions).[14] However, measuring the effects of physiotherapy interventions on these outcomes can be challenging.

The first challenge that researchers face in using patient-relevant outcomes in trials is how to measure them adequately. In 2005, the COnsensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) initiative was founded by an international multidisciplinary team of researchers, which caused an improvement in the development and evaluation of outcome measurement instruments, mainly patient-reported outcome measures (PROMs).[19] This advancement allowed researchers to use valid, reliable and responsive PROMs in their trials.

The second challenge is using this type of outcome as a primary outcome in randomised trials. PROMs commonly require a larger sample size and longer study duration to allow the research questions to be adequately powered and addressed.[14] Unfortunately, many trials of physiotherapy interventions have insufficient power, mainly due to small sample sizes and/or short periods of follow-up. Authors commonly emphasise secondary results instead of correctly addressing the main research of a randomised trial.[11,20]

Evaluating the effects of an intervention on PROMs with adequate internal validity and power can face important barriers, such as high research costs.[3,14] A possible solution is the use of surrogate endpoints (biomarker or intermediate outcomes), which are a substitute for target outcomes of interest in trials. A surrogate endpoint does not measure the clinical benefit of primary interest; rather, it is expected to predict that clinical benefit or harm based on epidemiological, therapeutic, pathophysiological or other scientific evidence. Surrogate endpoints are used to improve trial efficiency by reducing the sample size and shortening the duration of follow-up, thereby lowering the overall research costs;[3] these reductions, in some cases, may influence trial feasibility.[21,22] When a target outcome requires long follow-up, which is common in rare diseases, it is ethically feasible to use surrogate endpoints as a primary outcome in trials.[23] Furthermore, surrogate endpoints in small feasibility and pilot trials can guide the design and investment in a fully powered randomised trial planned to verify the effect of the intervention on the target outcome.[3]

Randomised trials that evaluate treatment effects using surrogate endpoints in physiotherapy have been developed and published.[24,25] As observed in other fields,[3] surrogate endpoints in physiotherapy trials have traditionally focused on biomarkers, such as carotid intima-media thickness, which is a surrogate endpoint of generalised atherosclerosis.[26] In addition, intermediate outcomes related to measures of symptoms (which can sometimes be regarded as surrogate endpoints) and function (which can be considered target outcome)[3] have also been used in physiotherapy, such as exercise capacity.[24] Exercise capacity can be measured directly by a cardiorespiratory assessment of peak oxygen uptake (VO_{2peak}) or indirectly by a submaximal test, such as the 6-minute walk test.[25] Exercise capacity has been used as a surrogate outcome for mortality, hospitalisation and health-related quality of life,[25] which are the final patient-relevant outcomes. On the other hand, exercise capacity can be also a target outcome if it is something that interferes with activities of daily living, especially if the patient reports his/her perception of effort, as assessed by the Borg Scale.[27] Four criteria were established by the recently published

https://doi.org/10.1016/j.jphys.2024.11.013

framework for classifying an intermediate outcome as a surrogate or a target endpoint: type of measurement; whether the outcome is patient-reported or represents health benefits; intention of the intervention or hypothesis of the trial; and association with the target outcome.[3] Researchers must apply those criteria for the correct use of intermediate outcomes.[3]

Despite the advantages provided by surrogate endpoints, they increase the uncertainty of the true value of the intervention. In this context, it is important to ensure that the observed changes in a surrogate endpoint are sufficiently reliable to predict the efficacy of the intervention in the patient-relevant target outcome. It is therefore recommended that the use of surrogate endpoints in randomised trials be limited to those that have been validated. Several levels of evidence must be provided for a surrogate endpoint to be considered valid: biological or clinical plausibility of the relationship between the surrogate endpoint and the target outcome (Level 3); consistent association between the surrogate endpoint and the target outcome, provided by a series of observational or epidemiological studies (Level 2); and association between the treatment effect on the surrogate endpoint and the treatment effect on the target outcome, preferably provided by a meta-analysis of randomised trials (Level 1).[25,28] Therefore, the highest level of evidence is recommended for the use of a surrogate endpoint in trials: evidence of a strong association between the intervention effect on the surrogate endpoint and the intervention effect on the patient-relevant target outcome.[3,6,7,28]

Let us consider our example of whether exercise capacity, as a surrogate outcome measured by the gold standard approach (maximal or symptom-limited test that provides the value of VO_{2peak}) or by indirect submaximal test[25] with adequate criterion-validity with the gold standard approach (the 6-minute walk test),[29] is valid for mortality, hospitalisation and health-related quality of life in heart failure for exercise rehabilitation interventions. Exercise capacity is the maximum amount of physical exertion that a person can sustain[30] and the relationship between exercise and mortality, disease and health has long been recognised[31] (Level 3). Epidemiological studies have shown that an increase in VO_{2peak} of 3.5 ml/kg/min (1.0 metabolic equivalent) corresponds to a risk reduction in mortality in individuals with cardiovascular disease of 12%, including individuals with heart failure[31] (Level 2). A meta-analysis of 31 randomised trials of exercise-based cardiac rehabilitation compared with no exercise control (involving 4,784 participants with heart failure) showed a low level of association between improvements in exercise capacity (VO_{2peak} or 6-minute walk test) and mortality and hospitalisation. In the same meta-analysis, a moderate association between exercise capacity and health-related quality of life was found. Therefore, exercise capacity, measured directly (VO_{2peak}) or indirectly (6-minute walk test), has poor validity as a surrogate outcome for mortality and hospitalisation and a moderate validity as a surrogate outcome for health-related quality of life in people with heart failure submitted to exercise-based cardiac rehabilitation (Level 1).[25] Therefore, this scientific evidence supports the use of exercise capacity as a surrogate outcome for health-related quality of life in trials with a similar intervention and population.[25]

Unfortunately, researchers commonly use surrogate outcomes that are invalid for predicting the effect in the patient-relevant outcome.[23,32] According to a previous review, less than half of the included randomised trials that used a surrogate outcome as the primary outcome discussed its validity to predict the effect of the intervention in the patient-relevant outcomes. The use of surrogate outcomes that are invalid can be misleading and result in the implementation of harmful interventions.[32] For example, in rheumatology, the biomarker serum urate (SU) has traditionally been used and recommended by The Outcome Measures in Rheumatology (OMERACT) Gout Working Group as a surrogate outcome. Gout is a common inflammatory arthritis that manifests as episodes of intense swelling, joint pain and disability. In studies involving participants with gout, SU was used as a substitute for clinically meaningful and patient-relevant outcomes such as reduction in the number of flares,[33,34] functional disability and health-related quality of life.[33] However, there is limited evidence that changes in SU or achieving target SU (reduction of SU to < 6 mg/dl) are associated with changes in patient-relevant outcomes in gout. In 2018, a meta-regression analysis that was performed with data from 10 RCTs did not confirm an association between SU and gout flare, except in studies with longer duration. The authors pointed out the biological plausibility of the relationship between the SU and the management of gout, but highlighted that SU did not reach the required level of evidence to be considered a surrogate outcome of patient-relevant outcomes in gout.[34] In 2021, the OMERACT Gout Working Group published a critical review aimed at outlining the potential of SU to be a surrogate outcome and the challenges for confirming the validity of SU as a surrogate in gout trials,[33] even for flare, which is the most investigated patient-relevant outcome. The role of SU as a surrogate of other patient-relevant outcomes in gout (such as pain, functional disability and health-related quality of life) is even less clear.[33]

Some commonly investigated outcomes in the physiotherapy field, such as grip strength and inflammatory biomarkers, have the potential to be surrogate endpoints for many patient-relevant outcomes. Grip strength has been identified as a potential surrogate endpoint for global muscle strength in myopathy, motor neuron disease[35] and stroke,[36] as well as for all-cause mortality in the healthy population.[37] It is considered an important and reliable biomarker of health, considering its consistent association in explaining function, fracture susceptibility, nutritional deficiency, cognitive decline, depression and overall quality of life,[38,39] mainly in the elderly population.[38] Therefore, it was recommended to routinely assess grip strength in clinical practice throughout the lifespan.[38,39] In chronic pain, inflammatory biomarkers have been associated with a decrease in pain, fatigue, physical function and disability, and an increase in mental health and health-related quality of life in subjects with chronic low back pain, knee osteoarthritis, fibromyalgia and chronic fatigue syndrome.[18] However, to be used as surrogate outcomes in future RCTs, the validity of grip strength and inflammatory biomarkers as surrogates for these patient-centred outcomes must be carefully investigated for each health condition as well as for each patient-relevant outcome. Evidence of a strong association between the intervention effect on the surrogate endpoint and the intervention effect on the patient-relevant target outcome (Level 1 of evidence) is recommended.[6,7,28,32] Unfortunately, this was not observed for grip strength or for inflammatory biomarkers for the previously mentioned health conditions and patient-relevant outcomes.

Validation of surrogate outcomes is challenging. Two solutions were proposed to aid in surrogate validation: planning to collect data on both surrogate and target outcomes in future randomised trials, even in a subset of the sample, and making the data freely accessible for future validation. Finally, it is important to point out that the validity of surrogate endpoints is context-specific. If a surrogate endpoint has been validated for one patient-relevant outcome or for one health condition, it is not automatically valid for other patient-relevant outcomes or other health conditions.[32]

Aiming to improve the identification and definition of surrogate endpoints as well as the planning, design and reporting of trials that use surrogate endpoints, a definitional framework was recently proposed for surrogate endpoints,[3] and frameworks that extend the definition and interpretation of surrogate endpoints in trials were developed: the SPIRIT-Surrogate extension checklist[6] and the CONSORT-Surrogate extension checklist.[7] The development of the checklists followed the EQUATOR (Enhancing the QUAlity and Transparency Of health Research) guidelines.[6,7] In summary, the SPIRIT-Surrogate extension includes nine items modified from the SPIRIT 2013 checklist[6] and the CONSORT-Surrogate extension includes nine items modified from the CONSORT 2010 checklist and two new items[7] (Table 1). The items include explicit mention of using surrogate endpoints, providing evidence (or lack thereof) of surrogate endpoint validation, informing trial participants on the use of surrogate endpoints, and interpretation of findings in the context of using a surrogate endpoint.

Table 1
Summary of the SPIRIT- and CONSORT-Surrogate extension items.[6,7]

Main section of SPIRIT or CONSORT	Surrogate extension item	Checklist to which the extension item applies
Abstract/Trial summary	1a. State a) that the primary outcome is a surrogate endpoint, and b) the target outcome(s) whose intervention effect is being substituted for.	SPIRIT-Surrogate CONSORT-Surrogate
Introduction/Background/Objectives	1b. State a) that the primary outcome is a surrogate endpoint, and b) the target outcome(s) whose intervention effect is being substituted for.	SPIRIT-Surrogate CONSORT-Surrogate
Outcomes	2. State the practical or scientific reason(s) for using a surrogate endpoint as a primary outcome.	SPIRIT-Surrogate CONSORT-Surrogate
	3. State what other surrogate endpoints were considered and why the current one(s) were chosen.	SPIRIT-Surrogate
	4. Justification for selected surrogate endpoint: a) evidence (or lack thereof) of surrogate endpoint validation and b) evidence (or lack thereof) of validity being specific to the context used (eg, intervention, disease, population).	SPIRIT-Surrogate CONSORT-Surrogate
Sample size	5. Clarify if the sample size will be/was estimated to demonstrate that a minimum effect on the surrogate endpoint would be predictive of a benefit on the target outcome(s).	SPIRIT-Surrogate CONSORT-Surrogate
Ethics/Patient and public engagement	6. State whether and how trial participants will be/were engaged and informed before enrolment, and that the trial was designed to evaluate an intervention's effect using a surrogate endpoint.	SPIRIT-Surrogate CONSORT-Surrogate
Results/Outcomes and estimation	7. If the primary outcome is a composite outcome that includes a surrogate endpoint, report the intervention effect on all components.	CONSORT-Surrogate
Harms/Discussion	8. Comment on whether the trial design (including sample size and follow-up period), given the use of a surrogate endpoint, adequately captures the potential harms of the intervention being tested.	SPIRIT-Surrogate CONSORT-Surrogate
Statistical methods/Discussion	9. State what the plans are to conduct subsequent analyses/studies to verify current findings on the target outcome(s).	SPIRIT-Surrogate CONSORT-Surrogate
Discussion/Interpretation	10. Interpretation of findings of the trial in the context of using a surrogate primary endpoint, including its known validity for intervention effects on the target outcome and the potential benefit-risk assessments of the tested intervention for participants.	CONSORT-Surrogate
Dissemination/Data access	11. If surrogate and target outcome data will be/were collected in the trial, state the open access arrangements for the data for future secondary research.	SPIRIT-Surrogate CONSORT-Surrogate

Extensions should be used along with the main SPIRIT and CONSORT checklists. Extensions are applicable to all intervention trials using surrogate endpoints (based on any definition) as primary outcome(s). However, items could be relevant to report in non-randomised trials, observational studies and other studies using surrogate endpoints. Items can be combined or reported in different sections to the ones suggested in the extension. The specific item sections are recommendations and not requirements. CONSORT = Consolidated Standards of Reporting Trials, SPIRIT = Standard Protocol Items: Recommendations for Interventional Trials.

Researchers, editors and reviewers should embrace a common definitional framework for surrogate endpoints[3] and then endorse those extensions[6,7] to provide clinicians and stakeholders with research results related to outcomes that are of interest and relevance to trial participants, patients, clinicians and other stakeholders.[3,6,7] Adherence to these extensions will enhance trial feasibility, transparency, interpretation, reporting and overall value, thus improving evidence-based practice and benefiting care provided by physiotherapists.

Competing interests: Nil.

Source(s) of support: The research was funded as part of the development of SPIRIT and CONSORT extensions by the UK Medical Research Council (grant number MR/V038400/1). CDCMF receives research productivity fellowships from the National Council for Scientific and Technological Development (CNPq/Brazil – Grant: 08516/2021-4).

Acknowledgements: Nil.

Correspondence: Christina DCM Faria, Department of Physiotherapy, Escola de Educação Física, Fisioterapia e Terapia Ocupacional, Universidade Federal de Minas Gerais, Minas Gerais, Brazil. Email: cdcmf@ufmg.br or chrismoraisf@gmail.com

References

1. Chan AW, et al. *Rev Panam Salud Publica*. 2015;38:506–514.
2. Schulz KF, et al. *J Pharmacol Pharmacother*. 2010;1:100–107.
3. Ciani O, et al. *E Clin Med*. 2023;65:102283.
4. Kelly G, et al. *Physiotherapy*. 2018;104:359–366.
5. Elkins MR. *J Physiother*. 2015;61:103–105.
6. Manyara AM, et al. *BMJ*. 2024;386:e078525.
7. Manyara AM, et al. *BMJ*. 2024;386:e078524.
8. Gianola S, et al. *Phys Ther*. 2013;93:1456–1466.
9. Hoffmann TC, et al. *BMJ*. 2013;347:f3755.
10. Negrini S, et al. *Am J Phys Med Rehabil*. 2019;98:258–265.
11. Komukai K, et al. *Arch Phys Med Rehabil*. 2024;105:150–156.
12. da Silva TM, et al. *Man Ther*. 2015;20:388–401.
13. Costa LOP. *Br J Sports Med*. 2018;52:950–951.
14. Portney L. *Foundations of Clinical Research—Applications to Evidence-Based Practice*. 4th ed. Philadelphia: F. A. Davis Company; 2020.
15. García-Hermoso A, et al. *Arch Phys Med Rehabil*. 2018;99:2100–2113.e5.
16. Stamp L, et al. *Semin Arthritis Rheum*. 2018;48:293–301.
17. Morillon MB, et al. *Semin Arthritis Rheum*. 2021;51:1378–1385.
18. Kawi J, et al. *Biol Res Nurs*. 2016;18:147–159.
19. Mokkink LB, et al. *BMC Med Res Methodol*. 2006;6:2.
20. Jette AM, et al. *Phys Ther*. 2019;99:1119–1121.
21. Anand IS, et al. *J Am Coll Cardiol*. 2002;39:1414–1421.
22. Johnston K. *Neuroepidemiology*. 1999;18:167–173.
23. Manyara AM, et al. *J Clin Epidemiol*. 2023;160:83–99.
24. Gorman PH, et al. *Spinal Cord*. 2019;57:471–481.
25. Ciani O, et al. *JACC Heart Fail*. 2018;6:596–604.
26. Ghardashi-Afousi A, et al. *J Diabetes Complications*. 2020;34:107469.
27. Scherr J, et al. *Eur J Appl Physiol*. 2013;113:147–155.
28. Weir CJ, et al. *Pharm Stat*. 2022;21:740–756.
29. Pollentier B, et al. *Cardiopulm Phys Ther J*. 2010;21:13–21.
30. Goldstein RE. Chapter 8. Exercise Capacity. In: Walker HK, Hall WD, Hurst JW, eds. *Clinical Methods: The History, Physical, and Laboratory Examinations*. 3rd ed. Boston: Butterworths; 1990.
31. Kokkinos P, et al. *Circulation*. 2010;122:1637–1648.
32. la Cour JL, et al. *BMJ*. 2010;341:c3653.
33. Morillon MB, et al. *Semin Arthritis Rheum*. 2021;51:1378–1385.
34. Stamp L, et al. *Semin Arthritis Rheum*. 2018;48:293–301.
35. Smith B, et al. *Neuromuscul Disord*. 2023;33(S1):S170.
36. Martins JC, et al. *Braz J Phys Ther*. 2015;19:498–506.
37. García-Hermoso A, et al. *Arch Phys Med Rehabil*. 2018;99:2100–2113.e5.
38. Bohannon RW. *Clin Interv Aging*. 2019;14:1681–1691.
39. Vaishya R, et al. *J Health Popul Nutr*. 2024;43:7.

Websites

Physiotherapy Evidence Database www.PEDro.org.au
OMERACT Gout Working Group omeract-us.org/gout