

ORIGINAL ARTICLE

Clinical replicability of rehabilitation interventions in randomized controlled trials reported in main journals is inadequate

Stefano Negrini^{a,b}, Chiara Arienti^b, Joel Pollet^{b,*}, Julia Patrick Engkasan^c, Gerard E. Francisco^d, Walter R. Frontera^e, Silvia Galeri^b, Kamila Gworys^f, Jolanta Kujawa^f, Mazlina Mazlan^c, Farooq A. Rathore^g, Fabienne Schillebeeckx^h, Charlotte Kiekens^h, and the REREP study participants

^aDepartment of Clinical and Experimental Sciences, University of Brescia, Brescia, Italy

^bIRCCS Fondazione Don Carlo Gnocchi, Milan, Italy

^cDepartment of Rehabilitation Medicine, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia

^dThe Neuro-Recovery Research Center, TIRR Memorial Hermann, Houston, TX, USA

^eDepartment of Physical Medicine, Rehabilitation, and Sports Medicine, University of Puerto Rico School of Medicine, San Juan, Puerto Rico

^fDepartment of PRM, Medical University of Lodz, Lodz, Poland

^gDepartment of Rehabilitation Medicine, PNS Shifa Hospital, Karachi, Sindh, Pakistan

^hPhysical and Rehabilitation Medicine, University Hospitals, Leuven, Belgium – Department of Development and Regeneration, KU Leuven, Leuven, Belgium

Accepted 11 June 2019; Published online 18 June 2019

Abstract

Objective: The objective of this study was to study if randomized controlled trials (RCTs) in rehabilitation (a field where complex interventions prevail) published in main journals include all the details needed to replicate the intervention in clinical practice (clinical replicability).

Study Design and Setting: Forty-seven rehabilitation clinicians of 5 professions from 7 teams (Belgium, Italy, Malaysia, Pakistan, Poland, Puerto Rico, the USA) reviewed 76 RCTs published by main rehabilitation journals exploring 14 domains chosen through consensus and piloting.

Results: The response rate was 99%. Inter-rater agreement was moderate/good. All clinicians considered unanimously 12 (16%) RCTs clinically replicable and none not replicable. At least one “absent” information was found by all participants in 60 RCTs (79%), and by a minimum of 85% in the remaining 16 (21%). Information considered to be less well described (8–19% “perfect” information) included two providers (skills, experience) and two delivery (cautions, relationships) items. The best described (50–79% “perfect”) were the classic methodological items included in CONSORT (descending order: participants, materials, procedures, setting, and intervention).

Conclusion: Clinical replicability must be considered in RCTs reporting, particularly for complex interventions. Classical methodological checklists such as CONSORT are not enough, and also Template for Intervention Description and Clinical replication do not cover all the requirements. This study supports the need for field-specific checklists. © 2019 Elsevier Inc. All rights reserved.

Keywords: RCTs; Applicability; Complex interventions; Rehabilitation

Conflict of interest: None related to this work. S.N. is the Chief-Editor of the European Journal of Physical and Rehabilitation Medicine and Director of Cochrane Rehabilitation; W.R.F. is the Chief-Editor of the American Journal of Physical Medicine and Rehabilitation and President of the International Society of Physical and Rehabilitation Medicine; C.K. is the Coordinator of Cochrane Rehabilitation; J.P.E. and F.R. are in the Executive Committee of Cochrane Rehabilitation; C.A. and J.P. are part of the Headquarters of Cochrane Rehabilitation.

* Corresponding author. IRCCS Fondazione Don Carlo Gnocchi, Via Capececiaturo 66, 20148 Milan, Italy. Tel./fax: +390307245282.

E-mail address: joel.pollet.ft@gmail.com (J. Pollet).

1. Introduction

Clinical research should inform clinical practice by developing interventions that improve patient care. In this context, applicability (external validity or generalizability) has been defined as “the extent to which the effects observed in studies reflect the expected results of an intervention in ‘real-world’ conditions” [1], reproducibility as “the replication of results by reperforming the same analysis of the same data by a different analyst” [2], and

What is new?**Key findings**

- Only 16% of RCTs in rehabilitation (complex interventions) are considered applicable by clinicians.

What this adds to what was known?

- Concept of clinical replicability as “the accurate description in published reports of clinical studies of all details needed to apply the intervention in everyday clinical practice”.
- Classical RCTs checklists, such as CONSORT, are not sufficient for the reporting of complex interventions.

What is the implication and what should change now?

- In reporting checklists, a focus on clinical replicability is needed, to allow a better use by clinicians of the produced evidence.

replicability as “the replication of results by reperforming of the experiment collecting new data” [2]. These terms have been applied mainly to epidemiological and methodological research. In the present study, we focus, from a clinical perspective, on a specific aspect of applicability, which we call clinical replicability and define as “the accurate description in published reports of clinical studies of all details needed to apply the intervention in everyday clinical practice”. Clinical replicability is related to PICO elements (Patients, Interventions, Comparison, and Outcomes) [3].

Completeness of information about PICO elements, description of randomization, blinding and statistical analysis used, and the context of care, are usually ensured through the use of high-quality reporting checklists, such as CONSORT for RCTs [4] or STROBE for observational trials [5]. Unfortunately, while there are plenty of data on factors affecting the risk of bias, or internal validity, very few refer to applicability [1]. In some clinical situations, dosage and timing could be sufficient to describe the intervention (e.g., in pharmacological studies), but this may not apply to complex interventions (conventionally defined as interventions with several interacting components) or interventions delivered by multiprofessional teams. A checklist, the Template for Intervention Description and Clinical replication (TIDieR) [6], has been developed for all evaluative study designs to improve the reporting of interventions for replicability. A methodological guide to assess the applicability of comparative interventions studies [1] and specific checklists [7–10] have been developed, but that approach is limited to epidemiological studies.

Rehabilitation needs are steadily growing worldwide because of aging of the population, an increase in the prevalence of noncommunicable diseases, and a rise in the number of persons experiencing disability [11]. For this reason, the World Health Organization is developing strategies to encourage governments to scale up rehabilitation services worldwide [12]. Physical and rehabilitation medicine (PRM) is the medical specialty dealing with rehabilitation and could be a good example of a field where clinical replicability can be studied. Rehabilitation in fact uses a biopsychosocial approach [13], focuses mainly on functioning [14], and combines multiple interventions at different points in time through multiprofessional and interdisciplinary team work [13]. For these reasons, research in rehabilitation is particularly challenging [15].

Our hypothesis is that existing reporting checklists, including the CONSORT for nonpharmacological interventions [16] and the TIDieR [10], or those specifically developed for rehabilitation [17,18], do not address well enough the issue of clinical replicability. To test this hypothesis, we developed the REplicability of RCTs in Everyday rehabilitation clinical Practice (REREP) Study. We asked representative expert PRM teams around the world to evaluate a sample of RCTs recently published in the main PRM journals [19,20]. Our final aim was to verify if RCTs in rehabilitation include all the practical details needed of PICO elements to be able to clinically replicate the studied intervention in different clinical settings.

2. Methods**2.1. Design**

A survey of a predefined sample of PRM clinical expert teams chosen to be representative of the different areas of the world and clinical rehabilitation competencies.

2.2. Participants

Two authors (S.N., C.K.), who have worldwide knowledge of the PRM community, selected a convenience group of 10 teams including professionals from the 3 different areas of the world as defined by the International Society of PRM: 4 from Europe, Eastern-Mediterranean and Africa, 4 from Americas (2 North, 2 Central/South), and 2 from Asia and Oceania. The rehabilitation teams had to be multiprofessional; team leaders needed to have research experience but also an everyday clinical practice; team professionals had to be clinicians working in multiple clinical rehabilitation levels of care and specialties (acute, postacute, chronic; in/out-patients; musculoskeletal, neurological, cardiorespiratory, etc.). Seven teams from Belgium, Italy, Malaysia, Pakistan, Poland, Puerto Rico, and the USA agreed to participate. Three teams (Argentina, Jordan, and the USA) did not participate for the following reasons: 1 never answered, 2 agreed but 1 had to stop for unforeseen circumstances, and 1 did not submit results in due time.

The final sample included 47 individual participants: 20 PRM physicians (PRMp), 12 physiotherapists (PT), 6 occupational therapists (OT), 6 rehabilitation psychologists (PSY), and 3 others (OTH), including one psychiatrist and two speech and language therapists.

2.3. Selection of published randomized controlled trials

We selected all the RCTs published between January and June 2017 by the journals defined primary by the European Society of PRM according to specific criteria [19,20]. We included RCTs from American Journal of Physical Medicine and Rehabilitation (PM&R), Archives of

PM&R, Clinical Rehabilitation, Disability and Rehabilitation, European Journal of PRM, International Journal of Rehabilitation Research, Journal of Rehabilitation Medicine, PM&R. We excluded the online-only RCTs (not yet printed), and those consisting of secondary analysis. On December 7, 2017 we searched all these journals in PubMed using the date filter January 1, 2017 to June 30, 2017. Of 1,349 article we found 206 RCTs. Seventy-six were included after two independent reviewers (C.A., J.P.) checked the full articles, with a third investigator (S.N.) in case of disagreement for final decision. The reasons for exclusion are listed in the PRISMA diagram (Fig. 1).

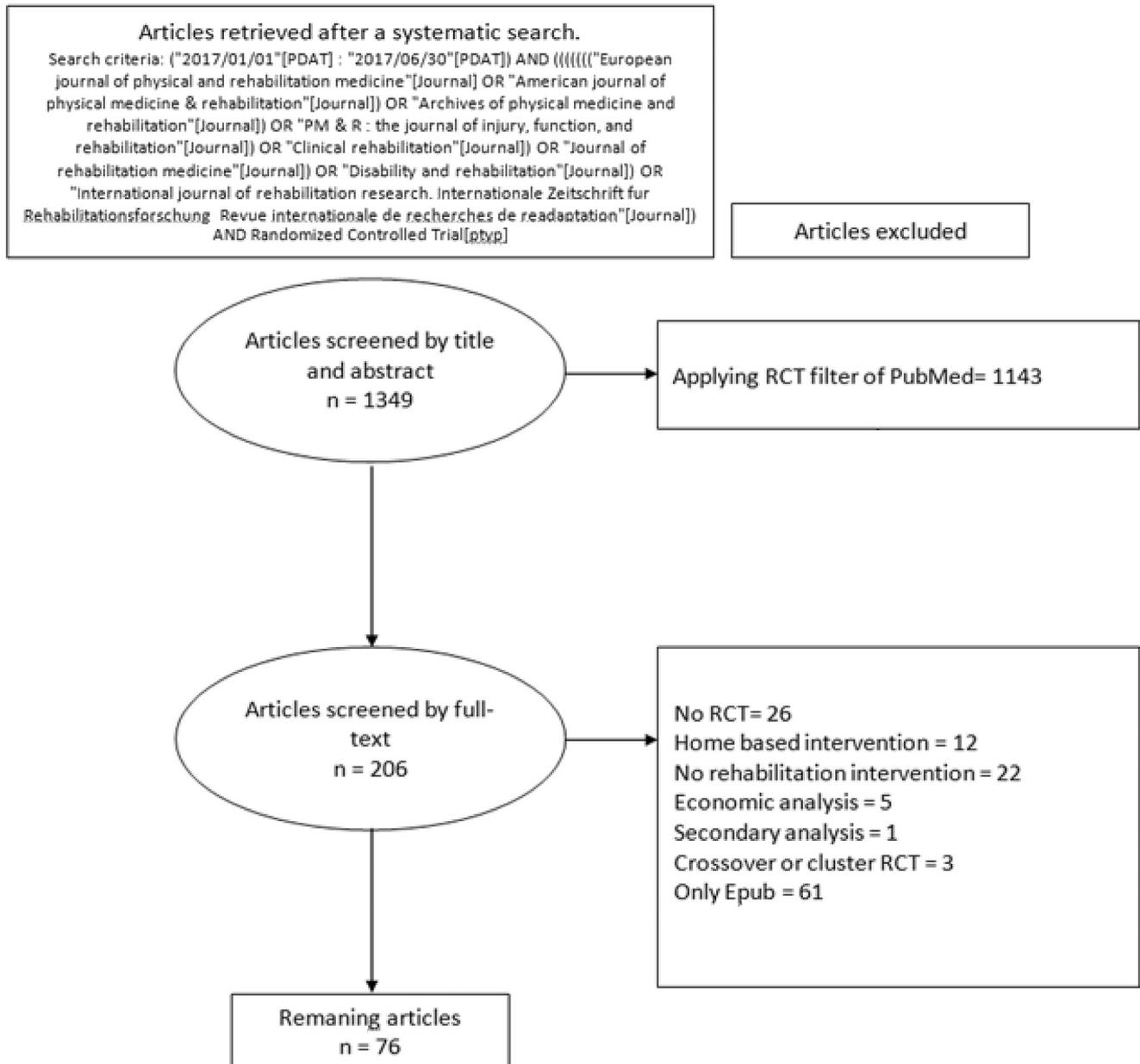


Fig. 1. PRISMA diagram of RCTs selection. Abbreviations: RCTs, randomized controlled trials.

Table 1. Contents of the questionnaire: items, possible answers to be provided per each item, and categorization of answers used for the two analysis performed

Questions		Answers to be provided per each item and subitem		
Item	Subitems	Questionnaire	Categorization 1	Categorization 2
Setting	Health care setting	Explicit information in	Perfect	Present
	Location	the “Methods” section		
Participants	Participants features	complete information	Imperfect	
	Adherence	partial information		
Intervention		other sections of the paper		
Materials		complete information		
Procedures		partial information		
Provider	Provider	Implicit information	Absent	Absent
	Skills	Absent		
	Experience	Not applicable		
Delivery	Cautions			
	Relationships			
	Intervention details			
	Order			
Clinical replicable?				

2.4. The survey

The survey focused on the intervention and not on the control group or the outcomes. We prepared a first draft of the survey from TIDieR [6], CONSORT [16] and the validity items developed for Benchmarking Controlled Trials [21]. After discussion among the main authors (S.N., C.A., J.P.) a consensus on the preliminary questionnaire was reached. The draft was piloted and amended during 2 meetings with the Italian team analyzing 10 RCTs (Appendix 1). A final consensus was reached.

The Survey is included in Appendix 2 and resumed in Table 1. It consists of 14 closed questions to 7 items: setting, participants, interventions, materials, procedures, providers, and delivery. A 15th key question “do you have enough information to replicate the intervention in your clinics?” was added.

2.5. Data collection

The main author (S.N.) had a Webex meeting with all the team leaders to inform them about the details of the study. The RCTs were distributed to the team leaders who decided whether the RCT was appropriate for their clinical reality (i.e., could be considered according to their setting, even if the type of treatment was not actually provided). The team leaders distributed the RCTs in their team according to expertise and area of clinical practice. From this moment, they did not have any role in data collection. Two authors (C.A., J.P.) had a Webex Meeting with each team to review the survey and answer questions. They remained available for methodological questions by email and/or Skype calls. Data have been collected individually

with a Survey Monkey filled in by each single clinician without any Consensus procedure into the different teams.

2.6. Data analysis

We categorized the answers according to Table 1.

We looked at the rate of agreement of each pair of responders and the resulting kappa has been judged as follows: <0.2 poor, 0.2–0.39 fair, 0.4–0.59 moderate, 0.6–0.79 good, 0.8–1 very good. The reliability has been checked within each profession and for the two analyses. We considered all the RCTs that had at least 4 filled surveys (9 paired comparisons) by responders of the same profession and we included a maximum of 5 RCTs per profession, starting from those with more answers. We had 5 RCTs for PRMp (8, 8, 8, 7, 7 responders, corresponding to 28, 28, 28, 21, 21 pairs, respectively) and PT (7, 6, 6, 6, 6 responders, corresponding to 21, 15, 15, 15, 15 pairs), and 3 for PSY (5, 4, 4 responders, corresponding to 10, 6, 6 pairs) and OT (4, 4, 4 responders, corresponding to 6, 6, 6 pairs).

To look for differences in the difficulty of clinical replication, we also analyzed results per profession using the chi-square test. Because teams included all the professionals involved in the clinical replicability in each specific context, we also performed an analysis per team, with the aim to check if the studies could be applied in each clinical reality as a whole. Finally, we performed a content analysis of the last open question to identify possible future improvement of the “clinical replicability questionnaire”.

2.7. Role of the funding source

The article has been produced without an external funder. Each professional participated on a voluntary basis. The

Table 2. Characteristics of participants and RCT analysis performed in the study

Characteristics of participants and RCT	Median	Minimum	Maximum
Participants recruited by each team	8	3	13
Number of participants per professions recruited by each team	3	1	5
RCT analyzed by each team according to its clinical competences	56	33	66
RCT analyzed by each profession	12	2	74
Number of participants analyzing each single RCT	9.5	3	16

institutes of the participating teams had no role in study design, data collection, analysis, interpretation, and report writing. All the authors had full access to all data and had final responsibility for the decision to submit for publication.

3. Results

The general characteristics of the participants are reported in Tables 2 and 3; Appendix 3 includes all included RCTs listed per topic area.

The response rate was 99%. The average rate of agreement was moderate (56%) and very good (80%) for the 3 (perfect, imperfect, absent) and 2 options (present, absent) analysis, respectively. Looking at the single items, we had 5 good, 9 moderate, and 1 fair average rate of agreement for the 3 options analysis and 9 very good, 5 good, and 1 moderate for the 2 options (Table 4). Owing to the pairs of participants contributing to the analysis, we also checked the distribution of obtained pair of agreements as reported in Table 5.

Overall, all participants considered unanimously that 12 (16%) RCTs were clinically replicable and none not replicable. Among the others, 56 (74%) RCTs have been considered replicable and 45 (59%) not replicable by at least one complete team. Looking at the single answers to our questionnaire, no study had all the information. At least one “absent” information was found by all participants in 60 RCTs (79%), and by a minimum of 85% in the remaining 16 (21%): 2 studies (3%) had 3 participants (29% of answers) who found at least one item not “absent”, 4 (5%)

had 2 (20% of answers), and 11 (14%) had 1 (10% of answers).

No Item reached 80% of “perfect” answers, whereas 4 were below 20% (Fig. 2). The topics with at least 50% of “perfect” answers included participant’s characteristics (79%), materials (65%), procedures (65%), order of interventions (63%), health care setting (63%), intervention (60%), and intervention details (59%). The answers with less than 20% included skills (19%), cautions (14%), experience (10%), and relationships (8%). The latter 4 topics were also the only topics with more than 50% of “absent”: 59%, 60%, 78%, and 80%, respectively. Apart from procedures (6%), health care setting (5%) and participant’s characteristics (1%), all the other topics had absent information in at least 9% of the cases.

There were no significant differences in the total number of answers according to profession: “perfect” ranged 36–47%, “absent” 25–30%. Looking at the single items, we did not find any difference among professions for health care setting, location, participants, adherence, provider, experience, and cautions. OT showed bigger problems than all the other professionals for intervention (“absent” 22% vs. a range 0–10%), materials (30% vs. 8–12%), and procedures (15% vs. 5–8%) ($P < 0.05$). PSY had fewer problems with intervention (0% vs. 8–22%) and skills (35% vs. 58–62%) ($P < 0.05$) (Fig. 3).

The open question did not suggest any new item for the questionnaire: we had 644 (92%) nothing is missing, 38 (5%) no answers, and 18 (3%) complaining the need to check other articles (protocol).

Table 3. Clinical competences of included teams and RCTs analyzed

	Musculoskeletal	Neurological	Cardiological	Pneumological	Pediatric	Geriatric	Other
RCTs (number)	26	36	3	2	1	5	3
Belgium	26 (100%)	28 (78%)	0	2 (100%)	1 (100%)	5 (100%)	2 (66%)
Italy	14 (54%)	19 (53%)	2 (66%)	1 (50%)	0	2 (40%)	0
Malaysia	21 (81%)	33 (92%)	2 (66%)	0	1 (100%)	5 (100%)	1 (33%)
Pakistan	15 (58%)	6 (17%)	0	0	0	1 (20%)	0
Poland	23 (88%)	30 (83%)	3 (100%)	1 (50%)	1 (100%)	5 (100%)	3 (100%)
Puerto Rico	22 (85%)	27 (75%)	0	0	0	5 (100%)	2 (66%)
USA	3 (12%)	21 (58%)	0	1 (50%)	1 (100%)	2 (40%)	1 (33%)

Table 4. Rate of agreement for each item of the questionnaire according to the two possible analysis considered

Items	Analysis with 3 items (perfect, imperfect, absent)	Analysis with 2 items (present, absent)
Total	56%	80%
Setting		
Health Care Setting	54%	95%
Location	50%	88%
Participants		
Participants Features	70%	95%
Adherence	35%	65%
Intervention	55%	84%
Materials	58%	93%
Procedures	60%	90%
Provider		
Provider	43%	71%
Skills	49%	59%
Experience	70%	71%
Delivery		
Cautions	51%	60%
Relationships	64%	67%
Intervention Details	57%	82%
Order	58%	80%
Clinical replicable?	64%	100%

Agreement of each items was verified into each profession, and included from a minimum of 6 pairs of responders (occupational therapists and psychologists) to a maximum of 28 (Physical and Rehabilitation Medicine physicians)—see main text for details. Percentages relate to mean agreement across participant pairs.

4. Discussion

This RREP study tested the hypothesis that the existing reporting checklists for RCTs do not solve the issue of clinical replicability in a clinical setting. In the present study, RCTs were considered not clinically replicable by as many as 31% of the expert participants. The specific areas in the [methods and materials](#) section that were rated the lowest included provider (skills and experience) and delivery items (cautions and relationships). On the contrary, best described areas were those included in the classical methodological checklists including participants characteristics, materials, procedures, health care setting, intervention, and 2 of the delivery items (order and intervention details).

The Agency for Healthcare Research and Quality developed a methods guide on “Assessing the applicability of Studies when comparing medical interventions” [1]. The guide states that the applicability of a study should be judged separately for different outcomes, it depends on context and cannot be assessed with a universal rating system, and it is best reported separately from the strength of the evidence. These recommendations, however, do not apply to our results showing that clinical replicability differs from applicability. The characteristics of individual studies listed in a PICOS framework by these guidelines

are relevant. However, those characteristics have been framed from the perspective of judging studies from a methodological point of view and not from the perspective of evaluating the descriptions provided in the published reports that are needed to apply treatments in clinics. This is perhaps a subtle difference, but it is also substantial.

A study similar to ours compared RCTs with observational trials using four orthopedic surgical procedures [22]. The authors considered 34 factors, grouped in 7 items: clinical characteristics of patients, setting and center, generic items selected for all interventions, items selected for minimally invasive or computer-assisted navigation procedures, blood loss, and postoperative pain management protocol. These items are similar but not fully comparable to ours because of the specific differences in setting [1] and the reduced complexity of the intervention. The lowest absolute results (<20% of prevalence of the item in the considered studies) were found for preoperative pain, deformity and comorbidities, centers’ surgical volume, information provided to patients, preoperative care, protocols for thromboprophylaxis, antibiotic prophylaxis, and postoperative pain management, characteristics of navigation system (open or closed), and blood loss. These results from orthopedic surgery studies confirm the problems inherent to protocol descriptions (like in rehabilitation), as well as our results for provider (skills and experience) and delivery items (cautions and relationships).

Another study [23] of knee arthroscopy used the benchmarking method, which we also considered while developing our survey [21] but approached the question from a methodological and an epidemiological perspective. The author considered 25 factors grouped in 5 items: selection of patients/population to the study, completeness of baseline data, process data, outcome data, and statistical analysis. The study showed deficiencies in reporting of baseline characteristics and adherence to interventions. The same author, using the same method, looked at all RCTs published in the British Medical Journal, the Journal of the American Medical Association, the Lancet, and the New England Journal of Medicine in the first 9 months of 2017 [24] and found deficits in describing patients’ path before randomization, health care settings, environmental factors, and cointerventions. Even though different methodologies were used, these articles confirm the deficiencies we found in some specific areas in our study.

The limitations of our study include the possible lack of representativeness of our sample of RCTs. However, there was a good correlation among the different teams from countries that are clearly different in terms of gross income, health services, culture, and rehabilitation approach and systems. In addition, the differences between different professionals’ answers were small. Although our results may not be applicable to other medical specialties and/or other complex interventions, they are similar to some of the findings reported by others [1]. We limited our sample of RCTs to the main PRM journals so we could study the specific competences of rehabilitation reviewers to identify clinical

Table 5. Average of rate of agreements obtained inside each single profession with the two possible analysis

Items	Analysis with 3 items					Analysis with 2 items				
	Poor	Fair	Moderate	Good	Very good	Poor	Fair	Moderate	Good	Very good
Total			69%	31%					44%	56%
Setting										
Health care setting		38%	38%	6%	19%			6%	6%	88%
Location	13%	19%	38%	25%	6%			19%	13%	69%
Participants										
Participants features	6%		38%	19%	38%			6%	6%	88%
Adherence	13%	56%	31%		0%		6%	50%	25%	19%
Intervention	6%	50%	13%		31%			19%	25%	56%
Materials	6%	19%	44%	13%	19%			6%	13%	81%
Procedures		25%	25%	38%	13%			13%	13%	75%
Provider										
Provider	13%	44%	31%	6%	6%		13%	25%	25%	38%
Skills	6%	38%	19%	38%	0%		13%	38%	44%	6%
Experience		25%	19%	19%	38%		13%	31%	19%	38%
Delivery										
Cautions	13%	31%	25%	19%	13%		19%	44%	19%	19%
Relationships	6%	25%	13%	31%	25%		19%	19%	38%	25%
Intervention Details	6%	19%	44%	13%	19%			31%	13%	56%
Order	6%	38%	13%	25%	19%		13%	13%	19%	56%
Clinical replicable?	25%	38%	6%	31%					100%	

replicability. However, more general journals present similar problems [24]. Finally, in the analysis at 3 items, we had a low agreement among responders. In our view, this is not highly relevant because the low agreement could be explained by individual different approaches to where the information should be provided in the article.

As a group of clinicians, including three Editors-in-Chief of clinical rehabilitation journals, we felt the need to introduce in this study the concept of clinical replicability,

defined as “the accurate description in published reports of clinical studies of all details needed to apply the intervention in everyday clinical practice”. The clinical replicability could be used to identify the quality of description of PICO, which represents the main elements of the real-world evidence [3] and it could allow a better applicability of the intervention in different clinical settings. Therefore, clinical replicability is not reproducibility or replicability because it does not refer to research but to clinics. It is not

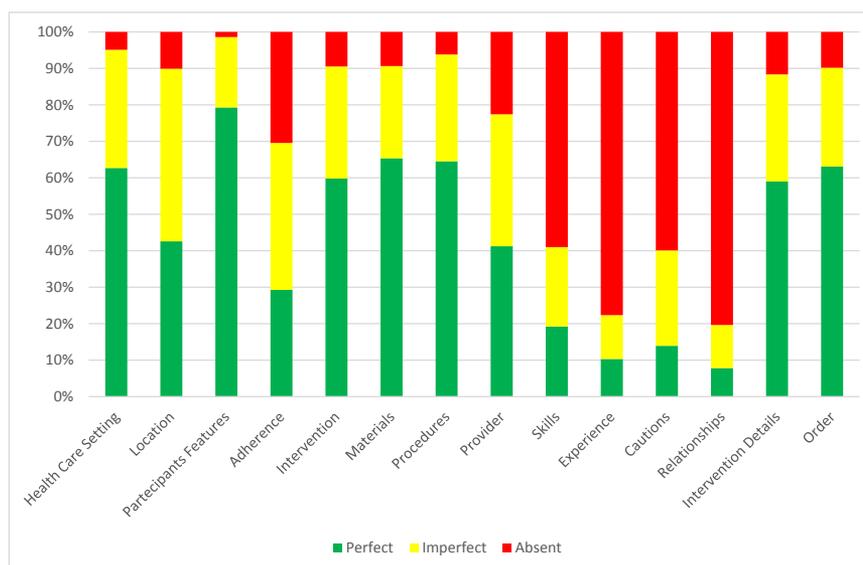


Fig. 2. Answers (percentage of TRIALS) to each single item of the questionnaire. The classical methodological items present in the CONSORT checklists have the best clinical replicability, while the worst correspond to the technical description of the intervention.

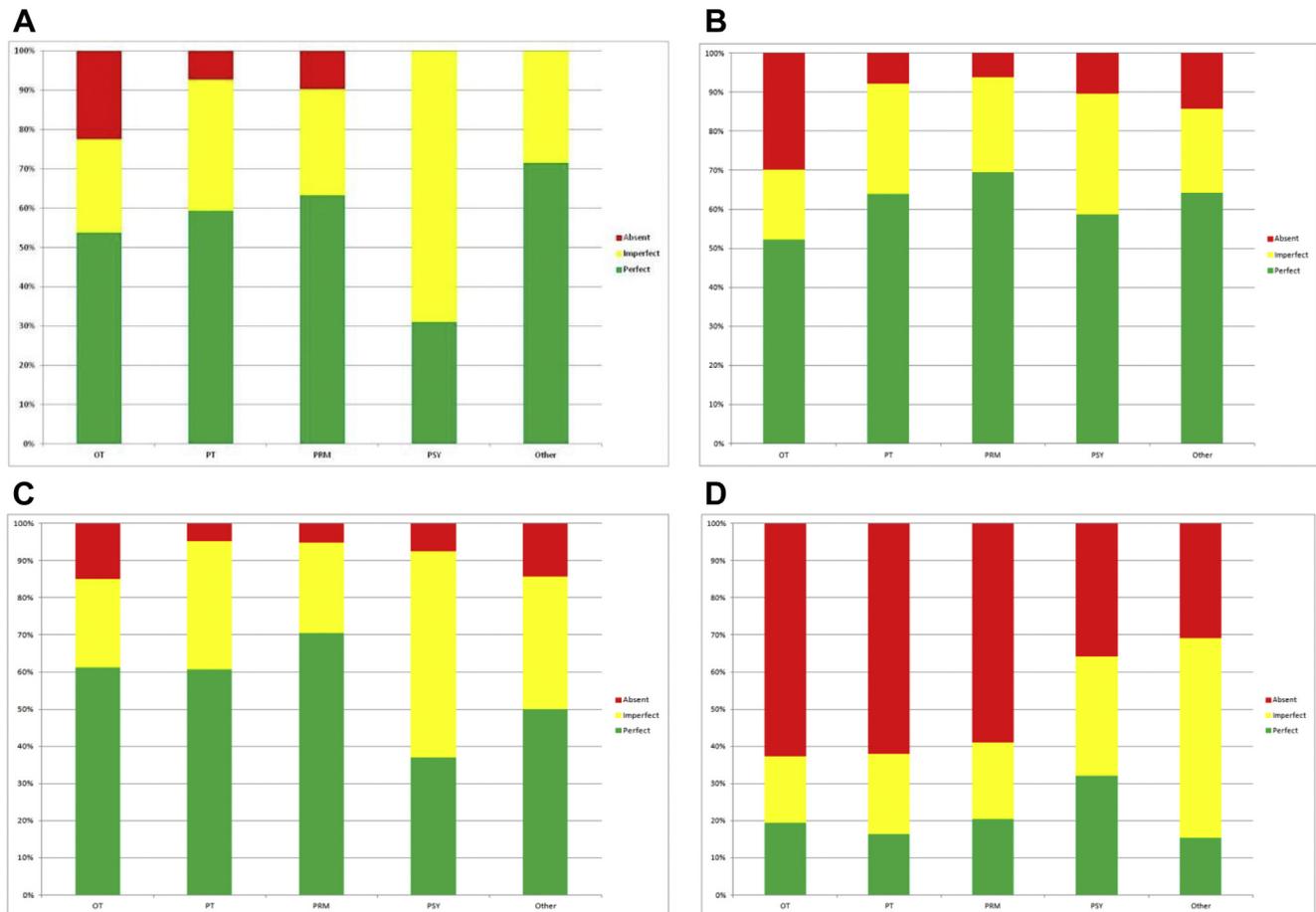


Fig. 3. Answers to the questionnaire for different professionals: (A) intervention; (B) materials; (C) procedures; (D) skills. Statistically significant differences from the other professionals have been shown with difficulties of OTs for intervention, materials, and procedures, while PSY showed less problems with intervention and skills. *Abbreviations:* OT, occupational therapists (41 RCT analyzed); PT, physical therapists (72 RCT analyzed); PRMp, PRM physicians (74 RCT analyzed); PSY, rehabilitation psychologists (11 RCT analyzed); OTH, others (14 RCT analyzed).

applicability (external validity or generalizability), that is much wider and include methodological and conducting issues, even if it could be interpreted as the part referring to reporting. Clinical replicability is not a single reporting issue, but a concept that should inform reporting. Conceptualizing clinical replicability as an independent issue could serve epidemiologist and methodologist to better focus the needs of clinicians. A good example of a checklist developed for clinical replicability is the TIDieR [6]. Although this informed part of the present study, it proved to be incomplete for rehabilitation research. In our opinion, clinical replicability should be highly context-specific and should inform specific checklists such as those for rehabilitation [17,18].

5. Conclusions

This study shows that there are problems in the clinical replicability of published RCTs in everyday clinical practice when complex interventions as exemplified by rehabilitation interventions are tested. These problems can be identified by all rehabilitation professionals involved in the team. The

areas with the least problems of clinical replicability are those generally better described by classic methodological checklists such as CONSORT. Conversely, more significant problems were found in the topics related to the human factors (typical of rehabilitation, but not only) such as skills, experience, and relationships. Unfortunately, also the item “cautions” revealed important problems. The results of the present study suggest the need for specific guidelines to improve clinical replicability of RCTs in rehabilitation. This specific case-study should be extended to other fields to verify the generalizability of our results.

CRedit authorship contribution statement

Stefano Negrini: Conceptualization, Formal analysis, Supervision, Methodology, Writing - original draft, Writing - review & editing. **Chiara Arienti:** Data curation, Formal analysis, Investigation, Writing - review & editing. **Joel Pollet:** Data curation, Formal analysis, Investigation, Writing - review & editing. **Julia Patrick Engkasan:** Supervision, Writing - review & editing. **Gerard E. Francisco:**

Supervision, Writing - review & editing. **Walter R. Frontera**: Supervision, Writing - review & editing. **Silvia Galeri**: Supervision, Writing - review & editing. **Kamila Gworys**: Supervision, Writing - review & editing. **Jolanta Kujawa**: Supervision, Writing - review & editing. **Mazlina Mazlan**: Supervision, Writing - review & editing. **Farooq A. Rathore**: Supervision, Writing - review & editing. **Fabienne Schillebeeckx**: Supervision, Writing - review & editing. **Charlotte Kiekens**: Supervision, Writing - review & editing.

Acknowledgments

The authors acknowledge the role of all the REREP Study participants. The authors wish to acknowledge the “Royal Society Te Apārangi, New Zealand”, for the support given to the study.

We acknowledge the role of all the REREP Study participants: University Hospitals Leuven, Physical and Rehabilitation Medicine, Leuven, Belgium; Coordinators: Charlotte Kiekens MD, Fabienne Schillebeeckx MD; Team: Stijn Bogaerts MD, Sofie Rummens MD, An De Groef PT, Geert Verheyden PT, Ann Spriet PT, Dieter Van Assche PT, Margreet Van Dijk PT, Bart Vrijssen PT, Blanche Staes OT, Griet Van Kerschaver OT, Eline Note PSY. IRCCS Don Gnocchi Foundation, Center E. Spalenza, Rovato, Italy; Coordinator Silvia Galeri MD; Team: Elisabetta Del Zotto MD, Luca Medici PT, Paolo Patelli PT, Alessandra Redolfi PSY. Department of Rehabilitation Medicine, University of Malaya, Kuala Lumpur, Malaysia; Coordinators: Mazlina Mazlan MD, Julia Patrick Engkasan MD; Team: Anwar Suhaimi MD, Chung Tze Yang MD, Aishah Ahmad Fauzi MD, Lee Poh Chen PT, Parimalanganthi Veradan PT, Noor Shahaneem Shaikh Mazran OT, Saufiyah Adnan OT. Department of Rehabilitation Medicine, PNS Shifa Hospital, DHA II, Karachi, Pakistan; Coordinator: Farooq Azan Rathore MD, Team: Shazia Farrukh PT, Asma Qamar OT. Department of PRM, Medical University of Lodz, Lodz, Poland; Coordinators: Jolanta Kujawa MD, Kamila Gworys MD; Team: Anna Puzder MD, Elżbieta Kowalewska MD, Jowita Gasztych PT, Marta Chrzanowska-Rydz PT, Justyna Redlicka OT, Magdalena Tomczak SLT, Katarzyna Binder PSY. Department of Physical Medicine, Rehabilitation, and Sports Medicine, University of Puerto Rico School of Medicine, San Juan, Puerto Rico; Coordinator: Walter Frontera MD, PhD; Team: William Micheo MD, Luis Baerga MD, Fernando Sepulveda MD, Carmen Lopez MD, Eduardo Ramos MD, Luis Cotto MD, Veronica Rodriguez MD. TIRR Memorial Hermann, Houston, USA; Coordinator Gerard E. Francisco MD; Team: Jason Chen MD, Radha Korupolu MD, Dharmendra Kumar MD, Marcia Kern PT, Erin E. Edenfield PT.

Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jclinepi.2019.06.008>.

References

- [1] Atkins D, Chang SM, Gartlehner G, Buckley DI, Whitlock EP, Berliner E, et al. Assessing applicability when comparing medical interventions: AHRQ and the Effective Health Care Program. *J Clin Epidemiol* 2011;64:1198–207.
- [2] Stevens JR. Replicability and reproducibility in comparative psychology. *Front Psychol* 2017;8:862.
- [3] Guyatt GH, Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, Schünemann HJ, GRADE Working Group. What is “quality of evidence” and why is it important to clinicians? *BMJ* 2008;336:995–8.
- [4] Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. *Ann Intern Med* 2010;152:726–32.
- [5] von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, et al. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med* 2007;147:573–7.
- [6] Hoffmann T, Glasziou P, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ* 2014;348:g1687.
- [7] Rothwell PM. External validity of randomised controlled trials: “to whom do the results of this trial apply?” *Lancet* 2005;365:82–93.
- [8] Bornhöft G, Maxon-Bergemann S, Wolf U, Kienle GS, Michalsen A, Vollmar HC, et al. Checklist for the qualitative evaluation of clinical studies with particular focus on external validity and model validity. *BMC Med Res Methodol* 2006;6:56.
- [9] Green LW, Glasgow RE. Evaluating the relevance, generalization, and applicability of research: issues in external validation and translation methodology. *Eval Health Prof* 2006;29(1):126–53.
- [10] Pibouleau L, Boutron I, Reeves BC, Nizard R, Ravaud P. Applicability and generalisability of published results of randomised controlled trials and non-randomised studies evaluating four orthopaedic procedures: methodological systematic review. *BMJ* 2009;339:b4538.
- [11] World Health Organization. Rehabilitation 2030: a call for action. Concept note. Available at www.who.int/disabilities/care/ConceptNote.pdf?ua=1. Accessed December 5, 2018.
- [12] World Health Organization. Rehabilitation 2030: a call for action. The need to scale up rehabilitation, 7/02/2019. Available at <https://www.who.int/disabilities/care/Need-to-scale-up-rehab-July2018.pdf?ua=1>.
- [13] European Physical and Rehabilitation Medicine Bodies Alliance. White book on physical and rehabilitation medicine in Europe. Introduction, executive summary, and methodology. *Eur J Phys Rehabil Med* 2018;54(2):125–55.
- [14] European Physical and Rehabilitation Medicine Bodies Alliance. White book on physical and rehabilitation medicine (PRM) in Europe. Chapter 3. A primary medical specialty: the fundamentals of PRM. *Eur J Phys Rehabil Med* 2018;54(2):177–85.
- [15] Negrini S. Evidence in Rehabilitation Medicine: between facts and prejudices. *Am J Phys Med Rehabil* 2019;98:88–96.
- [16] Boutron I, Altman DG, Moher D, Schulz KF, Ravaud P, CONSORT NPT Group. CONSORT statement for randomized trials of nonpharmacologic treatments: a 2017 update and a CONSORT extension for nonpharmacologic trial abstracts. *Ann Intern Med* 2017;167:40–7.
- [17] Page P, Hoogenboom B, Voight M. Improving the reporting of therapeutic exercise interventions in rehabilitation research. *Int J Sports Phys Ther* 2017;12(2):297–304.
- [18] Groeneweg R, Rubinstein SM, Oostendorp RA, Ostelo RW, van Tulder MW. Guideline for reporting interventions on spinal manipulative therapy: consensus on interventions reporting criteria list for spinal manipulative therapy (CIRCLe SMT). *J Manipulative Physiol Ther* 2017;40(2):61–70.

- [19] Franchignoni F, Stucki G, Muñoz Lasa S, Fialka-Moser V, Vanderstraeten G, Quittan M, et al. Publishing in physical and rehabilitation medicine: a European point of view. *J Rehabil Med* 2008; 40(6):492–4.
- [20] Franchignoni F, Ozçakar L, Michail X, Vanderstraeten G, Christodoulou N, Frischknecht R. Publishing in physical and rehabilitation medicine. An update on the European point of view. *Eur J Phys Rehabil Med* 2013;49(5):711–4.
- [21] Malmivaara A. Assessing validity of observational intervention studies –the Benchmarking controlled trials. *Ann Med* 2016;48: 440–3.
- [22] Pibouleau L, Boutron I, Reeves BC, Nizard R, Ravaud P. Applicability and generalisability of published results of randomised controlled trials and non-randomised studies evaluating four orthopaedic procedures: methodological systematic review. *BMJ* 2009; 339:b4538.
- [23] Malmivaara A. Validity and generalizability of findings of randomized controlled trials on arthroscopic partial meniscectomy of the knee. *Scand J Med Sci Sports* 2018;28(9):1970–81.
- [24] Malmivaara A. Generalizability of findings from randomized controlled trials is limited in the leading general medical journals. *J Clin Epidemiol* 2019;107:36–41.