ACRM
AMERICAN CONGRESS OF REHABILITATION MEDICINE

Archives of Physical Medicine and Rehabilitation

journal homepage: www.archives-pmr.org

Archives of Physical Medicine and Rehabilitation 2018;■:■■■-■■■

REVIEW ARTICLE

On the Reporting of Experimental and Control Therapies in Stroke Rehabilitation Trials: A Systematic Review

Keith R. Lohse, PhD,[a,b,c] Anupriya Pathania, MS,[a,b,c] Rebecca Wegman, BS,[c] Lara A. Boyd, PT, PhD,[d] Catherine E. Lang, PT, PhD[e]

From the [a]Department of Health, Kinesiology, and Recreation and [b]Department of Physical Therapy and Athletic Training, University of Utah, Salt Lake City, UT; [c]School of Kinesiology, Auburn University, Auburn, AL; [d]Department of Physical Therapy, University of British Columbia, Vancouver, British Columbia, Canada; and [e]Program in Physical Therapy, Program in Occupational Therapy, Department of Neurology, Washington University School of Medicine, St. Louis, MO.

Abstract

Objective: To use the *C*entralized *O*pen-*A*ccess *R*ehabilitation database for *S*troke to explore reporting of both experimental and control interventions in randomized controlled trials for stroke rehabilitation (including upper and lower extremity therapies).

Data Sources: The Centralized Open-Access Rehabilitation database for Stroke was created from a search of MEDLINE, Embase, Cochrane Central Register of Controlled Trials, Cochrane Database of Systematic Reviews, and Cumulative Index of Nursing and Allied Health from the earliest available date to May 31, 2014.

Study Selection: A total of 2892 titles were reduced to 514 that were screened by full text. This screening left 215 randomized controlled trials in the database (489 independent groups representing 12,847 patients).

Data Extraction: Using a mixture of qualitative and quantitative methods, we performed a text-based analysis of how the procedures of experimental and control therapies were described. Experimental and control groups were rated by 2 independent coders according to the Template for Intervention Description and Replication criteria.

Data Synthesis: Linear mixed-effects regression with a random effect of study (groups nested within studies) showed that experimental groups had statistically more words in their procedures (mean, 271.8 words) than did control groups (mean, 154.8 words) ($P<.001$). Experimental groups had statistically more references in their procedures (mean, 1.60 references) than did control groups (mean, .82 references) ($P<.001$). Experimental groups also scored significantly higher on the total Template for Intervention Description and Replication checklist (mean score, 7.43 points) than did control groups (mean score, 5.23 points) ($P<.001$).

Conclusions: Control treatments in stroke motor rehabilitation trials are underdescribed relative to experimental treatments. These poor descriptions are especially problematic for "conventional" therapy control groups. Poor reporting is a threat to the internal validity and generalizability of clinical trial results. We recommend authors use preregistered protocols and established reporting criteria to improve transparency.

Archives of Physical Medicine and Rehabilitation 2018;■:■■■-■■■

A thorough and complete reporting of methods in clinical trials is essential not only for reproducibility of research but also for the clinical interpretation and implementation of experimental methods. Despite a general understanding of the necessity of reporting, there is significant research[1-3] suggesting that reporting in clinical trials is poor. To address this problem, stakeholders have developed numerous checklists and guidelines to improve the reporting of biomedical research.[4-6] Despite these guidelines, a problematic outcome of this research is the finding that reporting of nonpharmaceutical interventions is worse than the reporting of pharmaceutical interventions (which are, in themselves, also poorly reported).[2,3] This difference is in some ways understandable

PROSPERO Registry No.: CRD42014009010.
Disclosures: none.

0003-9993/18/$36 - see front matter © 2018 by the American Congress of Rehabilitation Medicine
https://doi.org/10.1016/j.apmr.2017.12.024

because the dose, timing, frequency, and pathway of treatment are easier to define in pharmaceutical trials where active ingredients are directly measurable. However, the mere recognition of this difference does not negate the problems that poor reporting poses to rehabilitation medicine. In physical and occupational therapy interventions, for instance, the active ingredient(s) and dosage of therapy are often not clear, dose-response curves are not easily quantified, and treatments are highly variable.[7,8]

Accurate and complete reporting is especially critical in *control* groups for several reasons. First, complete reporting for control groups establishes the internal validity of experimental findings (ie, interpreting a difference between experimental and control groups is contingent on the adequacy of the control). Second, complete reporting is required for comparing interventions across experiments. For instance, the difference between therapy A and control A might be contrasted with the difference between therapy B and control B. If reporting is poor, perceived differences in the efficacy of experimental treatments might actually reflect differences in control treatments. In a pharmaceutical intervention, control groups might receive identically administered placebos. In rehabilitation trials, however, control groups might be described as "conventional" therapy but, despite the same name, differ significantly in their frequency, intensity, timing, and type of therapy.[8]

The ambiguous use of the term "conventional" therapy creates considerable confusion in the literature. If all the details were adequately reported, readers could understand and contrast what occurred in different "conventional" therapies. However, previous researchers[9,10] have noted that this is not the case and many critical details of therapy are often missing, referred to as the "black box" of therapy. Although this lack of detail has been noted, it has not been quantified or formally analyzed. Thus, a first step toward addressing the underreporting of methodological details is to quantify the problem in the field of stroke rehabilitation.

In the present analysis, our objectives were (1) to characterize the reporting of important methodological details in both experimental and control arms of stroke rehabilitation trials across different types of participants, interventions, and outcomes; (2) to assess potential differences in reporting for experimental vs control groups; (3) to identify potential areas of weakness in reporting, which need to be addressed by collective action by research stakeholders (eg, authors, reviewers, editors, publishers); and finally (4) to repeat objectives 1 to 3 specifically in those interventions reported as conventional therapy. Focusing on this subset of interventions is an essential first step in eliminating the "black box" of therapy, illustrating the variation and ambiguity in interventions that are described as "conventional" therapy.

The present analysis is part of the systematic review we conducted to construct the *C*entralized *O*pen-*A*ccess *R*ehabilitation database for *S*troke (SCOAR),[8,11] which includes data from randomized controlled trials (RCTs) for upper and lower extremity therapies in adults with stroke. From SCOAR, we analyzed existing variables describing the type, frequency, duration, and overall dose of therapy. In addition, independent coders extracted descriptions of the experimental and control therapies from the RCTs and assessed all groups separately according to the Template for Intervention Description and Replication (TIDieR) checklist.[5] TIDieR is a 12-point checklist that assigns points on the basis of the criteria described in table 1. Although the TIDieR checklist has previously been used to describe physical and occupational therapy interventions,[12-14] this analysis represents the first attempt to use the TIDieR checklist to describe experimental and control therapies separately.

List of abbreviations:

PEDro	**Physiotherapy Evidence Database**
RCT	**randomized controlled trial**
SCOAR	**Centralized Open-Access Rehabilitation database for Stroke**
TIDieR	**Template for Intervention Description and Replication**

Methods

Selection of studies

The present analysis is part of SCOAR; thus, all RCTs met the same eligibility criteria as that of the systematic review[8] (for a full list of the included RCTs, see supplemental appendix S1, available online only at http://www.archives-pmr.org/). SCOAR was constructed from a systematic search of MEDLINE, Embase, Cochrane Central Register of Controlled Trials, Cochrane Database of Systematic Reviews, and Cumulative Index of Nursing and Allied Health from the earliest available date to May 2014 that identified 215 independent studies that focused on motor rehabilitation in adults with stroke, totaling 489 independent therapy groups. The summary statistics for these groups represent 12,847 patients in total. Details of the systematic review and construction of the database have been presented previously[8] (see also PROSPERO Registration No.: CRD42014009010). The PICO criteria[15] for the review were as follows:

1. *Population:* Human adults with motor impairment as a result of stroke (regardless of etiology or prior stroke).
2. *Intervention:* Any physical or occupational therapy intervention that required active movement on the part of the participant.
3. *Control:* All studies had to be RCTs, and studies were required to explicitly state random assignment to groups. For our analysis, we coded groups as "control" if they were identified as controls by the original authors. Alternatively, if a group received "conventional," "routine," "standard," or "usual" care without being specifically named as control, it was assumed that this was a control condition.
4. *Outcomes:* Only validated assessments of impairment or functional motor capacity that were administered by a clinician were extracted as outcomes (this excludes self-report measures, neuroimaging/-physiological measures, and study-specific kinematic/kinetic measures).

Extraction of therapy descriptions

To describe how therapy was delivered, we used extant variables in SCOAR (to represent time in therapy), created a therapy description coding template (to extract additional information on how therapy was delivered), and calculated word and reference counts (to estimate the amount of space in the article devoted to therapy descriptions). These measures are described below.

We used a number of methodological variables in SCOAR that were relevant to our descriptive analysis. These variables were as follows: (1) whether the therapy was experimental or control;

Table 1 Brief description of the TIDieR checklist used to assess different arms of RCTs

Item	Description
1. Brief name	A short name or phrase (used by the authors) to describe the intervention is provided.
2. Why	The rationale, theory, or goal motivating the intervention is provided.
3. What — materials	Physical and/or informational materials used in the intervention is provided.
4. What — procedures	The sequence, activities, and/or processes used in the intervention is explained.
5. Provided by whom	A description of the person who provided the intervention (expertise, background, or specific training) is given.
6. How	A description of the modes of delivery of the intervention is provided.
7. Where	A description of where the therapy occurred (and any relevant infrastructure) is provided.
8. When and how much	A sufficient description of the frequency, duration, intensity, or dose of the intervention is provided.
9. Tailoring	A description of how the intervention was individualized (if individualized) is provided.
10. Modifications	Any deviations from the intended protocol are noted and described.
11. How well — planned	Intervention adherence or fidelity is assessed, and strategies to maintain adherence are described.
12. How well — actual	The extent to which the therapy is delivered "as planned" is described.

NOTE. A full version of the TIDieR checklist[5] is available at www.equator-network.org.

(2) a short name for the therapy (based on the description by the original authors); (3) the total duration of therapy (in days); (4) the estimated total time scheduled for therapy (in hours); (5) the method of reporting the dose of therapy (as scheduled time, actual time, active time, or repetitions in therapy); and (6) the Physiotherapy Evidence Database (PEDro) scale score for assessing risk of bias.

In addition, we developed a therapy description template for extracting information relevant to the therapy description (supplemental appendix 2, available online only at http://www.archives-pmr.org/). For all groups, experimental and control, the form included items for who provided the intervention therapy and for whether additional physical therapy, occupational therapy, and speech/language therapy were explicitly stated as occurring outside the intervention. The form included several different variables assessing therapy dose. To estimate dose, hours per day, days per week, weeks of therapy, and total amount of therapy (in hours) were all coded separately. Extracting these variables separately allows flexible coding. For instance, an RCT might report total therapy hours and duration (leaving hours per day and days per week blank) whereas another RCT might report hours per day, days per week, and total duration (from which total hours of therapy can be inferred).

In addition, time in therapy was coded "0" if no assessment of time in therapy was made, a "1" if scheduled therapy time was reported, a "2" if actual therapy time was reported (ie, adherence had to be assessed to get a 2), a "3" if active therapy time was reported (ie, 30min of walking without breaks in a 30-min session), and a "4" if repetitions of therapy were reported. These codes always applied to the highest resolution of therapy engagement that was available. That is, if a study reported both the time scheduled for the therapy and the total number of repetitions for the whole therapy (ie, not just repetitions for a single exercise), it was coded as a "4." (Note that we adopted the authors' definitions of repetitions in each trial; if authors described their therapy in repetitions, it was counted as repetitions.)

Finally, we wanted to perform a more in-depth text-based analysis of how experimental and control therapies were being described. Lines of text were extracted from the Methods sections of the original articles (including tables/figures referenced within these sections). The Methods section was specifically chosen for its relevance; however, it should be noted that some details of interventions might have been reported elsewhere (eg, in the Introduction sections). A pair of independent student coders (supervised by K.R.L.) read the Procedures subsection of each Methods section (or the whole Methods section if no Procedures subsection was identified). Each sentence of these sections was coded as applying to the procedures of the control group, the experimental group, or both, and then entered into a separate text document (see https://github.com/keithlohse/SCOAR/tree/master/group_desc). Coders were initially trained on a subset of 10 studies for consistent coding. After training, each student coded half of the studies in SCOAR and a senior author (K.R.L.) double-checked each student's coding (revising the coding until consensus was reached, if necessary). From these excerpts, we calculated word counts and reference counts for each group in each study.

Assessment of TIDieR criteria

TIDieR criteria are a more detailed measure of the quality/completeness of reporting that complement the other measures. Each group (experimental or control) was scored on all 12 dimensions of the TIDieR assessment and given a composite TIDieR score (with high values indicating better quality of reporting). Unlike the word and reference counts (which drew only from the Methods sections), TIDieR scores were based on the entire article. In cases where trials had published protocols,[16-23] the TIDieR score included information from both the protocol and the RCT. All TIDieR criteria were assessed by 2 independent coders (A.P. and R.W.). Coders read background material pertaining to TIDieR[5,12] and were trained on an initial subset of 10 RCTs from SCOAR for consistency. After training, coders independently rated all experimental and control interventions on the basis of the 12 TIDieR criteria summarized in table 1. Comparisons between coders were done incrementally, in batches of studies per week. In each batch, initial agreement between coders was high (94% across all batches) and any disagreement between coders was resolved through discussion with a senior author (K.R.L.).

Risk of bias assessment

Consistent with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement on systematic reviews, a

Table 2 Descriptive statistics for the reporting of experimental and control therapies

Variable	Experimental Groups (n=287)	Control Groups (n=202)
PT/OT/SLT outside the trial		
PT	74 (26)	51 (25)
OT	54 (19)	39 (19)
SLT	13 (5)	8 (4)
Frequency of therapy		
Hours per day	1.34±1.28 (0.17–8)	1.15±0.87 (0–6)
Hours per day reported as range	12 (4)	8 (4)
Hours per day missing	73 (25)	71 (35)
Days per week	4.16±1.05 (2–7)	4.24±1.14 (0–7)
Days per week reported as range	3 (1)	3 (1)
Days per week missing	83 (29)	76 (37)
Duration of therapy (wk)	6.13±5.10 (2–52)	6.31±5.74 (2–52)
Duration reported as range	0 (0)	0 (0)
Duration missing	49 (17)	49 (24)
Method of reporting dose of therapy		
0. None	53 (18)	46 (23)
1. Time scheduled	181 (63)	126 (62)
2. Actual time	14 (5)	12 (6)
3. Active time	24 (8)	13 (6)
4. Repetitions	15 (5)	5 (2)

NOTE. Values are mean ± SD (range) or n (%).
Abbreviations: OT, occupational therapy; PT, physical therapy; SLT, speech/language therapy.

risk of bias assessment was performed using PEDro scale scores included in SCOAR. Details of this assessment are provided in supplemental appendix 3 (available online only at http://www.archives-pmr.org/), but most studies scored between 6 and 8 points (out of 10) on the PEDro scale and were thus considered to be of good quality.

Statistical analysis

To control for the multilevel nature of our data (ie, experimental and control groups are nested within RCTs and thus likely to be statistically dependent), we used linear mixed-effects regression with a random effect of study identification number as our principal method of analysis.[24,25] All statistical analyses were performed in R[26] using the dplyr[27] and lme4[28] packages. Figures were generated using the ggplot2[29] package. Code and data for all analyses are available at https://github.com/keithlohse/SCOAR/tree/master/group_desc. For word counts, reference counts, and TIDieR criteria, a model with only a random effect of study ID number (ie, the "random intercept" model) was compared with a model with a random effect of study ID number and a fixed effect of group (experimental vs control) to test the fixed effect of group. Statistical significance was determined by a reduction in the Akaike information criterion and the *P* value for the group coefficient in the model. Thus, a reduction in the Akaike information criterion and a *P* value of <.05 would mean that including group as a factor explained significantly more variance than would be expected given the null hypothesis. (Because of the multilevel nature of the models, degrees of freedom and *P* values were estimated using the Satterthwaite approximation.[30])

In line with our specific aims, these analyses are presented in 2 sections below. The focus in the first section is on inferential statistics and differences in reporting for control groups vs experimental groups. The second section recreates those analyses specifically in the subset of groups that were described as "standard," "customary," "conventional," and/or "usual" care. This subset was mostly composed of control groups, but it is possible that experimental groups could be conventional therapy groups as well (for instance, if conventional therapy was administered at an experimental dose). Thus, the first section of the Results section shows areas of strength and weakness in the reporting of interventions *overall*. The second section of the Results section shows the variation (or the lack of detail) in treatment parameters for interventions that are all described as *conventional therapy*.

Results

Reporting of experimental vs control therapies

Descriptive statistics for the content of different therapies are presented in table 2.

Text analysis of therapy descriptions

Linear mixed-effects regression controlling for groups nested within studies revealed that experimental groups had significantly more words dedicated to the description of their procedures (271.8±158.8 words) than did control groups (154.8±112.6 words) ($t_{301.9}=12.27$; $P<.001$). (Descriptive statistics are reported as mean ± SD.) The difference in word counts is shown in figure 1A. Similarly, experimental groups had significantly more references in the description of their procedures (1.60±2.20) than did control groups (0.82±1.93) ($t_{281.3}=7.62$; $P<.001$). The difference in reference counts is shown in figure 1B.

Analysis of TIDieR criteria

With respect to TIDieR criteria, experimental groups scored significantly higher on the total TIDieR score (7.43±1.57 points)

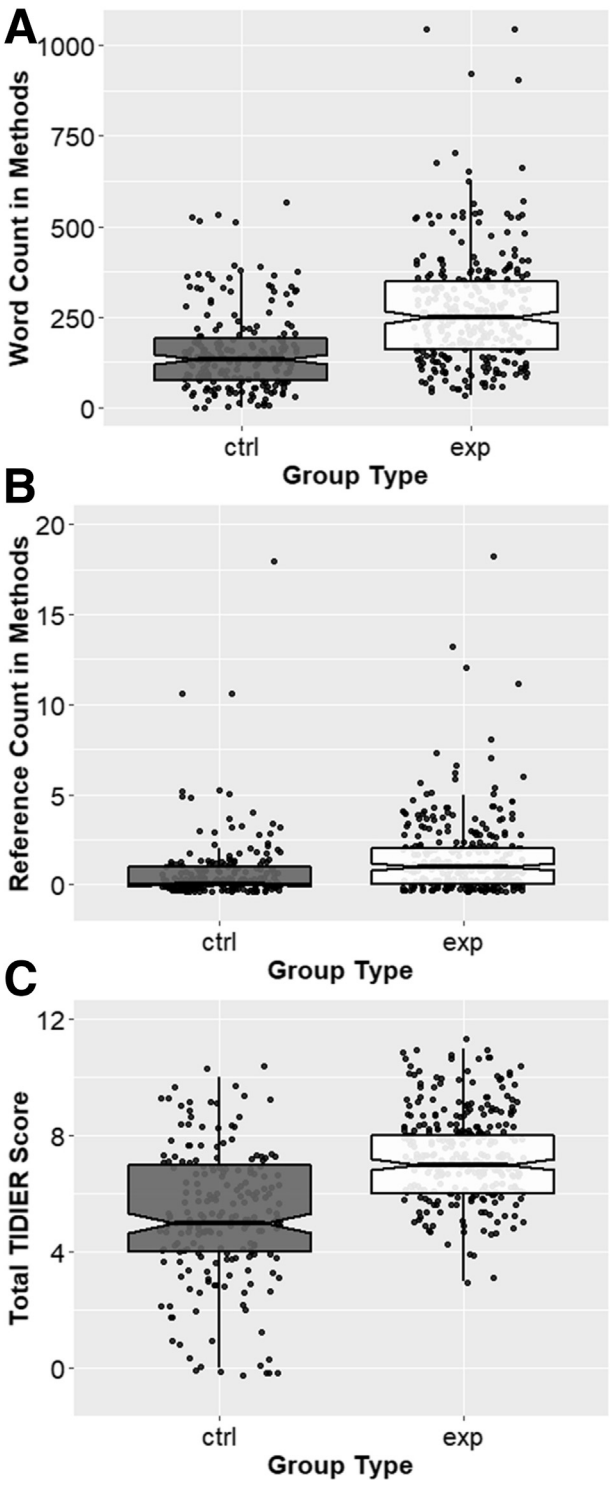

Fig 1 Word counts (A), reference counts (B), and total TIDieR scores (C) as a function of group type. Box plots superimpose the medians and interquartile ranges over the raw data. For graphing purposes, data points have been "jittered" on the x-axis so that all points are visible. Abbreviations: ctrl, control group; exp, experimental group.

than did control groups (5.23±2.25 points) ($t_{313.2}=15.38$; $P<.001$), as shown in figure 1C. Given this significant difference in overall TIDieR scores, we also performed exploratory analyses for the individual TIDieR criteria as a function of group. Comparing these individual dimensions of the TIDieR scale is important because it more clearly shows specific areas of strength/weakness in reporting. Because of the binomial outcome measure (ie, achieving the TIDieR criterion or not), these analyses used mixed-effects logistic regression, with the TIDieR criterion being predicted by a fixed effect of group. These data are presented in table 3.

We also performed exploratory analyses to see if total TIDieR scores changed over time and if TIDieR scores were correlated between experimental and control groups. These analyses allow us to see if there was a general trend for reporting to improve over time and if the quality of reporting was consistent within a study (ie, was better experimental group reporting associated with better control group reporting). With respect to time, linear mixed-effects regression revealed no reliable relation between the year of publication and the total TIDieR score ($t_{212.4}=.80$; $P=.48$). As shown in figure 2A, there was considerable variation in the quality of reporting; and controlling for year of publication, the difference between experimental and control groups remained statistically significant ($t_{313.2}=15.37$; $P<.001$).

For internal consistency, there was a significant positive relation between the TIDieR score for the experimental group and the TIDieR score for the control group. The regression coefficient (β) was .17, ($t_{244.6}=3.80$; $P<.001$), suggesting that articles with higher TIDieR scores for the experimental group also tended to show relatively higher TIDieR scores for the control group. However, as shown in figure 2B, there are several studies that had good reporting for the experimental group but poor reporting for the control group, which attenuated this relation.

It is also important to examine the relation between the overall TIDieR scores and the word/reference counts because these factors could be independent of each other. For example, one author could efficiently meet the TIDieR criteria across an article with a limited number of words in the Methods section; however, another author could spend considerable space giving details of a piece apparatus, and although such details are important, they would only meet limited TIDieR criteria (ie, *Materials* and possibly *Procedures*). As shown in figure 3A, there was a statistically significant positive relation between the word count and the overall TIDieR score, controlling for group (β=.005; $t_{471.0}=8.16$; $P<.001$). As shown in figure 3B, the relation between the TIDieR score and the reference count was positive but not statistically significant, controlling for group (β=.08; $t_{398.2}=1.76$; $P=.08$).

Reporting of "conventional" therapies

On the basis of the descriptions provided by the authors, we restricted our data to those therapies that were described as conventional, standard, usual, or customary care. This included conventional therapy that was dosage matched to an experimental intervention and thus might have been delivered at higher dose, but the contents were still conventional therapy. However, this subset excluded therapies that were described as conventional therapy *plus* a complimentary intervention (eg, conventional therapy plus "sham robotic therapy" was excluded). Of 109 total conventional therapy groups, 104 were identified as control groups and 5 were identified as experimental groups in which conventional therapy was delivered at a higher volume than normal. Descriptive statistics for all these conventional therapy groups are presented in table 4.

Table 3 Individual TIDieR criteria as a function of group type (experimental or control)

Criterion	Experimental Groups (n = 287)	Control Groups (n = 202)	*Z* Observed for the Effect of Group	*P*
1. Name	255 (89)	146 (72)	7.46	<.001
2. Why	277 (97)	89 (44)	5.25	<.001
3. What — materials	163 (57)	32 (16)	7.34	<.001
4. What — procedures	200 (70)	98 (49)	5.42	<.001
5. Administered by whom	213 (74)	112 (55)	7.54	<.001
6. How	199 (69)	120 (59)	3.81	<.001
7. Where	154 (53)	102 (50)	2.28	.02
8. When and how much	275 (96)	161 (80)	7.91	<.001
9. Tailoring	158 (55)	56 (28)	6.04	<.001
10. Modifications	7 (2)	5 (2)	1.01	.31
11. How well — planned	90 (31)	45 (22)	5.48	<.001
12. How well — actual	142 (49)	90 (45)	1.48	.14

NOTE. Values are n (%) for the group (control vs experimental) meeting the 12 TIDieR criteria. *Z* and *P* values come from mixed-effects logistic regression with a single fixed effect of group (control or experimental) and a random effect of study (as groups are nested within studies).

Discussion

Underreporting for control groups relative to experimental groups

In general, experimental groups had many more words (almost double) devoted to describing their procedures than did control groups. This pattern might be reasonable if the descriptions of control therapies were highly referenced, citing empirically based guidebooks, recommendations, or even internal documents for "best practices." However, control groups were also under-referenced, with 0.8 references on average, compared with experimental groups, with 1.6 references on average. An important limitation to this result is that these data are based only on text excerpted from the Methods sections of articles. Important information relevant to the procedures (especially things such as a theoretical rationale) could have been reported elsewhere (eg, in the Introduction sections).

The same pattern of poor reporting for control groups was also apparent in the much more comprehensive TIDieR scores. Control groups, on average, scored worse on the total TIDieR score than did experimental groups. Interestingly, there was also a positive relation between groups, such that better reporting for experimental groups was weakly associated with better reporting for control groups. When we break these results down into individual TIDieR criteria, we can see the strengths/weaknesses of reporting with much higher resolution.

Although experimental groups generally scored higher on TIDieR checklist than did control groups, there was a wide variation in the quality of reporting for both groups. As shown in table 3, for instance, experimental groups scored poorly for explaining *Where* therapy took place, the *Tailoring* of the intervention to individuals (or explicitly stating that the treatment was not individualized), any *Modifications* from the intended protocol (or explicitly stating that there were none), any plans to maintain adherence or assessment of the intervention as *Planned*, and, finally, experimental groups scored few points for explicitly describing the extent to which the therapy was *Actually* delivered as planned.

For control groups, more than half met TIDieR criteria for a short *Name* for the type of therapy administered, by *Whom* the therapy was administered, the general mode of *How* therapy was administered, and *When/How Much* therapy was administered. Control groups tended to score poorly on *Materials* and *Procedures*, which are essential to reproducibility of therapy and to establish the legitimacy of the control (ie, the internal validity of the comparison between the experimental group and the control group). Control groups also scored poorly for reporting *Tailoring* the protocol to the individual and/or *Modifying* the protocol.

Across both groups, *Tailoring*, *Modifications*, and *Planned/Actual* intervention adherence were poorly reported. Ambiguity in these areas is especially problematic because physical therapy, occupational therapy, and speech/language therapy are highly individualized in their delivery. This is not to say that it is easy to precisely report how therapy is individualized,[31] but it is essential to report how therapy is delivered as completely and efficiently as possible. However, to create optimal standards for reporting, there needs to be a better understanding of the "active ingredients" in therapy[32] and there needs to be considerable work done in developing effective tools for tracking how therapy is delivered. Recent work[33,34] on empirically validated taxonomies for therapy helps to address this problem, but further research is needed in this area. In the interim, it is advisable that authors provide sufficient reporting to meet all TIDieR criteria in their methods.

Interestingly, we did not see a significant trend for a change in the quality of reporting over time. This is perhaps not completely surprising as the TIDieR checklist was published in 2014 and our studies end in 2014,[5] but other standardized reporting guidelines predate the TIDieR checklist.[4,35] Furthermore, as shown in figure 2A, the reporting for experimental groups was significantly better than that for control groups even when controlling for year of publication. It is essential that stakeholders in the field work together to rectify this discrepancy, especially given that many of these control groups were collectively described as conventional therapy. Treating these control therapies as identical when they might be different undermines researchers' ability to perform meta-analyses and contrast results across trials.

Poor reporting of conventional therapy

When data were restricted to those therapies described as usual care or conventional therapy, we saw a wide variation in study parameters. For instance, studies were delivered anywhere

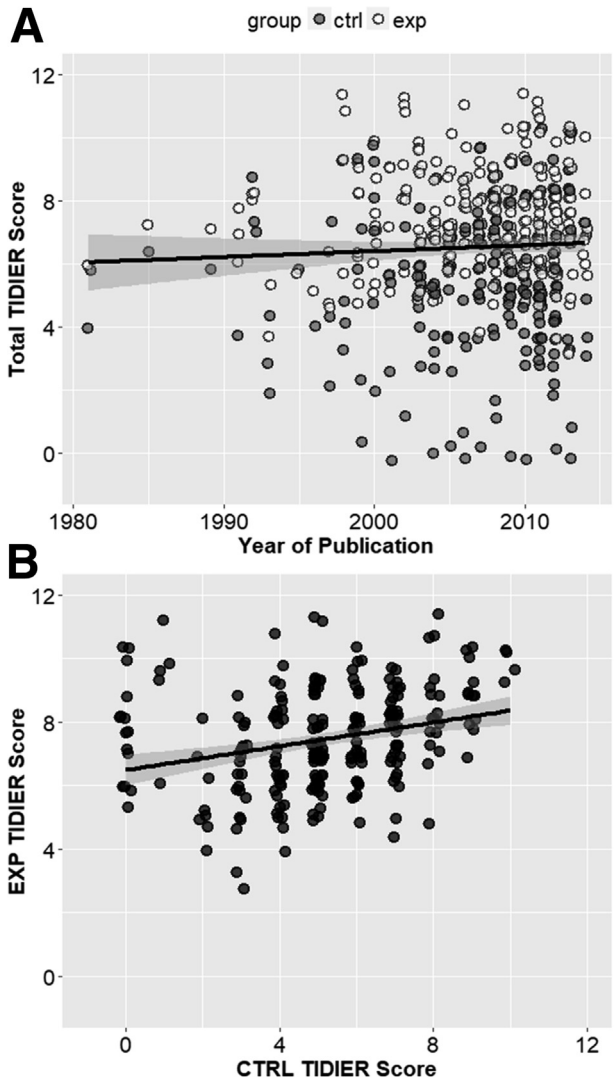

Fig 2 (A) Total TIDieR scores as a function of year of publication and group type. (B) Total TIDieR scores for experimental groups as a function of their corresponding control groups. For graphing purposes, TIDieR scores have been "jittered" so that all data points are visible. Lines show the line of best fit, and the shaded area shows the 95% confidence interval. Abbreviations: CTRL, control group; EXP, experimental group.

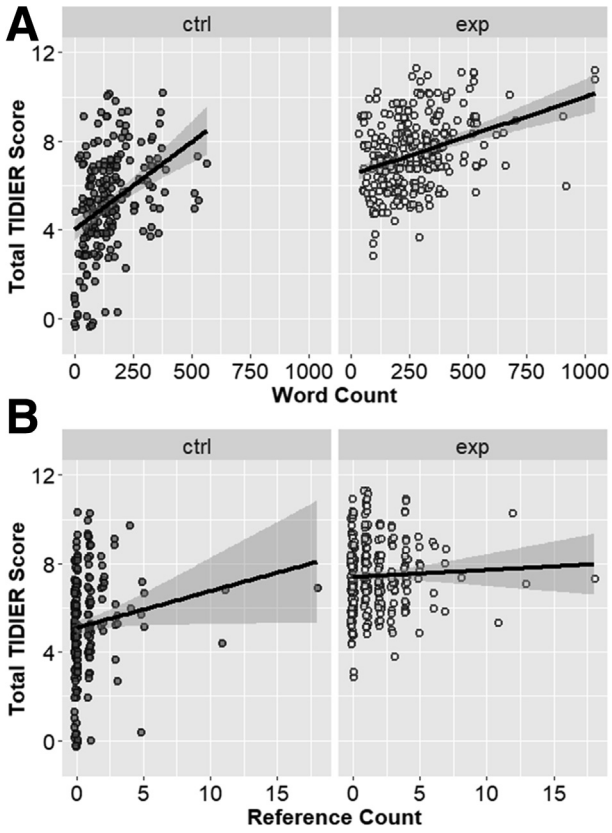

Fig 3 Total TIDieR scores as a function of word count (A) and reference count (B) in the Methods section of the original article. Data are plotted separately for the experimental and control groups. For graphing purposes, TIDieR scores have been "jittered" so that all data points are visible. Lines show the line of best fit in each group, and the shaded area shows the 95% confidence interval. Abbreviations: ctrl, control group; exp, experimental group.

between 2 and 26 weeks, from 3 to 7 days per week, with 0 to 4 hours of scheduled therapy per day. Some variation in these parameters is to be expected because conventional therapy will change from study to study on the basis of the study location, time poststroke, and different inclusion/exclusion criteria of individual studies. That said, this substantial variation is still problematic because readers will have a poor understanding of what happened during "conventional" therapies (see table 4).

Not only is there considerable variation in therapy parameters that are reported, but there is also considerable ambiguity in therapy parameters that are poorly reported. For instance, the theoretical rationale behind *Why* a therapy was chosen was seldom reported. We should note that this criterion was stringent in that points were not awarded for simply stating a theory behind the treatment (eg, stating Bobath-based, neurodevelopmental, or motor learning–based approach). Authors were required to clearly explain the rationale, theory, or goals of the therapy.[5] *Materials* and *Procedures* used during the intervention were also poorly reported. Although these elements are clearly critical, intervention materials are often missing (or not reported in sufficient detail) from intervention descriptions[36] and the included studies were no exception.

Finally, it was often not clear how/if the therapy was *Tailored* to the individual or generally *Modified* from the intended therapy in anyway. Although modifications might have been less frequent, it would be quite surprising if interventions were not tailored to the individual needs of participants in different trials. For instance, the TIDieR guidelines recommend a point for this criterion if the way in which an intervention was personalized, titrated, or adapted is sufficiently described.[5] Given large individual differences in functional capacity/ability, range of motion, age, and cognitive status in stroke trials, the current lack of reporting how therapy was tailored is problematic. Part of this lack of reporting might come from a lack of tools/methods for documenting how therapy is personalized over time. Researchers[37,38] have developed tools to track how therapy is administered to the upper extremity, but these methods (1) have not been widely adopted; (2) do not easily generalize beyond the upper extremity; and (3) remain to be validated as capturing the "active ingredients" of therapy. More work in this area is needed, and stroke rehabilitation would benefit from standardized tools to track how therapy is delivered over time. However, this statement must be balanced

Table 4 Descriptive statistics for the reporting of interventions described as "conventional" therapies (N = 109)

Variable	Value
PT/OT/SLT outside the trial	
PT	39 (36)
OT	26 (24)
SLT	6 (6)
Frequency of therapy	
Hours per day	1.22±0.74 (0.33–4)
Hours per day reported as range	6 (6)
Hours per day missing	28 (26)
Days per week	4.53±0.92 (3–7)
Days per week reported as range	3 (3)
Days per week missing	34 (31)
Duration of therapy (wk)	5.66±3.99 (2–26)
Duration reported as range	0 (0)
Duration missing	0 (0)
Method of reporting dose of therapy	
0. None	18 (17)
1. Time scheduled	74 (68)
2. Actual time	8 (7)
3. Active time	8 (7)
4. Repetitions	1 (1)
TIDieR criteria	
1. Name	87 (80)
2. Why	46 (42)
3. What — materials	12 (11)
4. What — procedures	48 (44)
5. Administered by whom	63 (58)
6. How	68 (62)
7. Where	55 (50)
8. When and how much	91 (83)
9. Tailoring	23 (21)
10. Modifications	3 (3)
11. How well — planned	24 (22)
12. How well — actual	56 (51)

NOTE. Values are mean ± SD (range) or n (%).
Abbreviations: OT, occupational therapy; PT, physical therapy; SLT, speech/language therapy.

against the fact that complete reporting is difficult and that the "active ingredients" that we need to be documenting are not clear.

Those points notwithstanding, authors can improve scientific communication by providing sufficient description for *all groups* to meet the 12 TIDieR criteria in their publications. Achieving these criteria would also improve the replicability of a given procedure. If more space is needed, authors should publish detailed trial protocols separately and/or use supplemental appendices to report details of therapy.

Study limitations

There are a number of limitations of the present review and analysis that need to be considered when interpreting these data. First, our analysis was restricted to the trials included in SCOAR, which currently goes on through 2014. The reason for this restriction is that this analysis was part of the original SCOAR systematic review. That said, these data come from 215 studies, totaling 489 independent treatment groups. Thus, although these data are not exhaustive, we think that this sample is representative of therapy descriptions in clinical trials. Similarly, there are many trials that were published after 2014 and not captured by this systematic review. However, when we analyzed TIDieR scores as a function of publication date, there was no evidence of a consistent trend for TIDieR scores to change over time.

Another concern is the quantification of time spent in therapy as a proxy for the "dose" of therapy. Although repetitions are the standard for tracking dose in preclinical animal trials[39,40] and in many human trials,[18,41,42] there are some areas of therapy in which repetitions may not be the most sensible criterion for tracking the delivery of therapy. For instance, if an occupational therapist is working on cooking skills with a client, there are many individual movements that might be included in a single "repetition" of a higher-order skill. Indeed, even in more constrained upper extremity exercises, the concept of a repetition has been defined in different ways.[18,42,43]

Conclusions

Overall, descriptions of control therapies were limited across RCTs. This lack of description is a problem for 2 reasons. First, poor reporting of control therapies hurts the internal validity of many RCTs, because it is impossible to establish the adequacy of the control without knowing in detail how materials and procedures were applied. Second, poor reporting hurts external validity, making it difficult to compare across trials and synthesize research. Currently, 2 studies could report using "conventional" therapy as a control, but those therapies could be substantially different. As a result, perceived differences in the efficacy of experimental treatments may be due to differences in the adequacy of the controls. To address these problems, we recommend that all details for all arms of a trial be reported sufficient to meet the 12 TIDieR criteria (publishing separate protocols or appendices as needed). Furthermore, there is a need to establish better process metrics, allowing researchers to more thoroughly document the administration of experimental and control treatments over time.

Keywords

Data reporting; Informatics; Rehabilitation; Research design; Stroke

Corresponding author

Keith R. Lohse, PhD, Department of Health, Kinesiology, and Recreation, University of Utah, 250 S 1850 E Room 200, Salt Lake City, UT 84112. *E-mail address:* rehabinformatics@gmail.com.

Acknowledgment

We thank Monica Nelson, BS, for her assistance in the text extraction and Emma Johnson, PhD, for her feedback on the statistical analyses.

References

1. Duff JM, Leather H, Walden EO, LaPlant KD, George TJ. Adequacy of published oncology randomized controlled trials to provide

therapeutic details needed for clinical application. J Natl Cancer Inst 2010;102:702-5.
2. Glasziou P, Meats E, Heneghan C, Shepperd S. What is missing from descriptions of treatment in trials and reviews? BMJ 2008; 336:1472-4.
3. Schroter S, Glasziou P, Heneghan C. Quality of descriptions of treatments: a review of published randomised controlled trials. BMJ Open 2012;2:e001978.
4. Schulz KF, Altman DG, Moher D, et al. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMC Med 2010;8:18.
5. Hoffmann TC, Glasziou PP, Boutron I, et al. Better reporting of interventions: Template for Intervention Description and Replication (TIDieR) checklist and guide. BMJ 2014;348:g1687.
6. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the PRISMA statement. Ann Intern Med 2009;151:264-9.
7. Lang CE, Lohse KR, Birkenmeier RL. Dose and timing in neurorehabilitation: prescribing motor therapy after stroke. Curr Opin Neurol 2015;28:549-55.
8. Lohse KR, Schaefer SY, Raikes AC, Boyd LA, Lang CE. Asking new questions with old data: the Centralized Open-Access Rehabilitation Database for Stroke. Front Neurol 2016;7:153.
9. DeJong G, Horn SD, Gassaway JA, Slavin MD, Dijkers MP. Toward a taxonomy of rehabilitation interventions: using an inductive approach to examine the "black box" of rehabilitation. Arch Phys Med Rehabil 2004;85:678-86.
10. DeJong G, Horn SD, Conroy B, Nichols D, Healton EB. Opening the black box of poststroke rehabilitation: stroke rehabilitation patients, processes, and outcomes. Arch Phys Med Rehabil 2005;86:1-7.
11. Mohabbati-Kalejahi N, Yazdi MAA, Megahed FM, et al. Streamlining science with structured data archives: insights from stroke rehabilitation. Scientometrics 2017;113:969-83.
12. Hoffmann TC, Walker MF. "TIDieR-ing up" the reporting of interventions in stroke research: the importance of knowing what is in the "black box". Int J Stroke 2015;10:657-8.
13. Yamato T, Maher C, Saragiotto B, et al. The TIDieR checklist will benefit the physical therapy profession. Phys Ther 2016;96:930-1.
14. van Vliet P, Hunter SM, Donaldson C, Pomeroy V. Using the TIDieR checklist to standardize the description of a functional strength training intervention for the upper limb after stroke. J Neurol Phys Ther 2016;40:203-8.
15. Schardt C, Adams MB, Owens T, Keitz S, Fontelo P. Utilization of the PICO framework to improve searching PubMed for clinical questions. BMC Med Inform Decis Mak 2007;7:16.
16. Duncan PW, Sullivan KJ, Behrman AL, et al. Body-weight-supported treadmill rehabilitation after stroke. N Engl J Med 2011;364:2026.
17. Harris JE, Eng JJ, Miller WC, Dawson AS. A self-administered Graded Repetitive Arm Supplementary Program (GRASP) improves arm function during inpatient stroke rehabilitation: a multi-site randomized controlled trial. Stroke 2009;40:2123-8.
18. Lo A, Guarino P, Richards L, Haselkorn J, Wittenberg G. Robot-assisted therapy for long-term upper-limb impairment after stroke. N Engl J Med 2010;362:1772-83.
19. Parry R, Lincoln N, Vass C. Effect of severity of arm impairment on response to additional physiotherapy early after stroke. Clin Rehabil 1999;13:187-98.
20. Saposnik G, Teasell R, Mamdani M, et al. Effectiveness of virtual reality using Wii gaming technology in stroke rehabilitation: a pilot randomized clinical trial and proof of principle. Stroke 2010;41:1477-84.
21. Dean CM, Ada L, Bampton J, Morris ME, Katrak PH, Potts S. Treadmill walking with body weight support in subacute non-ambulatory stroke improves walking capacity more than overground walking: a randomised trial. J Physiother 2010;56:97.
22. Timmermans AA, Verbunt JA, van Woerden R, Moennekens M, Pernot DH, Seelen HA. Effect of mental practice on the improvement of function and daily activity performance of the upper extremity in patients with subacute stroke: a randomized clinical trial. J Am Med Dir Assoc 2013;14:204-12.
23. van Delden AL, Peper CL, Nienhuys KN, Zijp NI, Beek PJ, Kwakkel G. Unilateral versus bilateral upper limb training after stroke: the upper limb training after stroke clinical trial. Stroke 2013;44:2613-6.
24. Snijders T, Bosker R. Multilevel analysis. Thousand Oaks: Sage; 2012.
25. Holmes Finch W, Bolin J, Kelley K. Multilevel modelling using R. Boca Raton: CRC Pr; 2014.
26. R Core Team. R: a language and environment for statistical computing. Available at: http://www.r-project.org/. Accessed September 17, 2014.
27. Wickham H, Francois R. dplyr: a grammar of data manipulation. R Package version 0.3.0.2. 2014. Available at: http://user2014.stat.ucla.edu/abstracts/talks/45_Wickham.pdf. Accessed August 17, 2015.
28. Bates D. lme4: mixed-effects modeling with R. New York: Springer; 2010.
29. Wickham H. ggplot2: elegant graphics for data analysis. New York: Springer-Verlag; 2009.
30. Kuznetsov A, Brockhoff PB, Christensen RHB. lmerTest: tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package), v. 2.0−25. 2013. Available at: http://cran.r-project.org/package=lmerTest. Accessed June 30, 2017.
31. Whyte J, Hart T. It's more than a black box, it's a Russian doll: defining rehabilitation treatments. J Phys Med Rehabil 2003;82:639-52.
32. Reinkensmeyer DJ, Burdet E, Casadio M, et al. Computational neurorehabilitation: modeling plasticity and learning to predict recovery. J Neuroeng Rehabil 2016;13:42.
33. Dijkers M, Hart T, Tsaousides T, Whyte J. Treatment taxonomy for rehabilitation: past, present, and prospects. Arch Phys 2014;95: S6-16.
34. Whyte J, Dijkers M, Hart T, Zanca J, Packel A. Development of a theory-driven rehabilitation treatment taxonomy: conceptual issues. Arch Phys 2014;95:S24-32.
35. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. Lancet 2001;357:1191-4.
36. Hoffmann T, Erueti C, Glasziou P. Poor description of non-pharmacological interventions: analysis of consecutive sample of randomised trials. BMJ 2013;347:f3755.
37. Donaldson C, Tallis R, Pomeroy V. A treatment schedule of conventional physical therapy provided to enhance upper limb sensorimotor recovery after stroke: expert criterion validity and intra-rater reliability. Physiotherapy 2009;95:110-9.
38. Hunter S, Crome P, Sim J, Donaldson C. Development of treatment schedules for research: a structured review to identify methodologies used and a worked example of 'mobilisation and tactile stimulation' for stroke patients. Physiotherapy 2006;92:195-207.
39. Kleim JA, Barbay S, Nudo RJ. Functional reorganization of the rat motor cortex following motor skill learning. J Neurophysiol 1998;80: 3321-5.
40. Nudo RJ, Wise BM, SiFuentes F, Milliken GW. Neural substrates for the effects of rehabilitative training on motor recovery after ischemic infarct. Science 1996;272:1791-4.
41. Lang CE, Macdonald JR, Reisman DS, et al. Observation of amounts of movement practice provided during stroke rehabilitation. Arch Phys Med Rehabil 2009;90:1692-8.
42. Lang C, Strube M, Bland M, et al. Dose-response of task-specific upper limb training in people at least 6 months post stroke: a phase II, single-blind, randomized, controlled trial. Ann Neurol 2016;80:342-54.
43. Stinear CM, Barber PA, Coxon JP, Fleming MK, Byblow WD. Priming the motor system enhances the effects of upper limb therapy in chronic stroke. Brain 2008;131:1381-90.