

# Where have all the pilot studies gone? A follow-up on 30 years of pilot studies in *Clinical Rehabilitation*

Clinical Rehabilitation  
2017, Vol. 31 (9) 1238–1248  
© The Author(s) 2017



Reprints and permissions:  
sagepub.co.uk/journalsPermissions.nav  
DOI: 10.1177/0269215517692129  
journals.sagepub.com/home/cre



Navaldeep Kaur<sup>1,2,3</sup>, Sabrina Figueiredo<sup>1,2,3</sup>,  
Vanessa Bouchard<sup>1,2,3</sup>, Carolina Moriello<sup>1,2</sup>  
and Nancy Mayo<sup>1,2,3</sup>

## Abstract

**Introduction:** Pilot studies are meritorious for determining the feasibility of a definitive clinical trial in terms of conduct and potential for efficacy, but their possible applications for planning a future trial are not always fully realized. The purpose of this review was to estimate the extent to which pilot/feasibility studies: (i) addressed needed objectives; (ii) led to definitive trials; and (iii) whether the subsequent undertaking of a definitive trial was influenced by the strength of the evidence of outcome improvement.

**Methods:** Trials published in the journal *Clinical Rehabilitation*, since its inception, were eligible if the word 'pilot' or 'feasibility' was specified somewhere in the article. A total of 191 studies were reviewed, results were summarized descriptively, and between-group effect sizes were computed.

**Results:** The specific purposes of piloting were stated in only 58% ( $n = 110$ ) of the studies. The most frequent purpose was to estimate the potential for efficacy (85%), followed by testing the feasibility of the intervention (60%). Only 12% of the studies were followed by a definitive trial; <4% of studies had a main study underway or a published study protocol. There was no relationship between observed effect size and follow-up of pilot studies, although the confidence intervals were very wide owing to small number of trials that followed on.

**Discussion:** Labelling and reporting of pilot studies needs to be improved to be concordant with the recently issued CONSORT guidelines. Feasibility needs to be fully tested and demonstrated prior to committing considerable human and monetary resources.

## Keywords

Rehabilitation, randomized controlled trial, controlled clinical trial

Received: 8 January 2017; accepted: 13 January 2017

<sup>1</sup>Division of Clinical Epidemiology, McGill University, Montreal, Quebec, Canada

<sup>2</sup>Center for Outcomes Research and Evaluation, McGill University Health Centre, Montreal, Quebec, Canada

<sup>3</sup>School of Physical and Occupational Therapy, McGill University, Montreal, Quebec, Canada

## Corresponding author:

Nancy Mayo, Center for Outcomes Research and Evaluation, Research Institute, McGill University Health Centre, 687 Pine Avenue West, Ross Pavilion, R4.17. Montreal, Quebec H3A 1A1, Canada.

Email: nancy.mayo@mcgill.ca

## Background

Clinical trials answer questions about deliberate interventions that are often innovations in treatment and the results are meant to inform clinical practice.<sup>1,2</sup> Because of the innovation, many of the details of a definitive trial are unknown before starting, and should be investigated systematically before committing to a larger study. Trials of rehabilitation interventions are particularly challenging as they often involve testing of interventions with different active ingredients.<sup>3,4</sup> How the multiple elements of the trial work together need to be tested, including the best way to identify participants, whether randomization is accepted, the processes around delivering the intervention, and the optimal control for the experimental intervention, to name but a few of the questions to be answered. Many trials suffer from recruitment challenges, high attrition rates because of tediousness or intolerable study demands<sup>2</sup> making it essential to identify these potential bumps in the trial road beforehand.

In the past, some authors have weighed in on definitions and purposes of pilot and/or feasibility studies.<sup>5-7</sup> The first consensus on the definitions of pilot and feasibility studies was published in 2016 and provided a conceptual framework to unify the disparate concepts that are grouped under feasibility or pilot studies.<sup>8</sup> Many pieces of research may need to be done before a main study can be justified, essentially to probe into the feasibility of various aspects of a study protocol. Important parameters, including recruitment rates, completion rates, adherence rates, and resources needed, that are crucial for designing a definitive trial, can be estimated through these preparatory studies.<sup>7</sup> However, if no evidence is provided to show that the intervention may produce some change in the outcome, with or without a control group, it is hard to judge the efficacy potential of an intervention. Both the processes planned for the trial and the potential for efficacy are necessary for a full trial to be feasible, but neither alone is sufficient.

The recently achieved consensus<sup>8</sup> is that ‘pilot’ studies fall under the dimension of ‘feasibility’ studies. Up to now, the terms ‘feasibility’ and ‘pilot’ have been used and misused in the medical

literature, particularly when these terms have been used ‘post hoc’ to disguise underpowered main studies,<sup>7,9</sup> those studies with methodological limitations, or those not completed because of inadequate funding.<sup>10</sup>

Feasibility of processes and outcome potential are essential elements for funding of a clinical trial, but the vast majority of feasibility studies are hardly published.<sup>4</sup> Those that are published tend to be the ones with data reporting efficacy potential. Some reviews on the methodological quality of pilot (feasibility) studies have been undertaken<sup>9,11-13</sup> and their conclusions are summarized in Table 1. Overall, it has been observed that the pilot studies are frequently employed for hypothesis testing and feasibility of processes are rarely considered. Although follow-up is recommended, only a meagre proportion of preliminary studies have been pursued further in a confirmatory study.

Over the last few years, a major emphasis has been laid upon informing researchers about appropriate objectives<sup>7,11,14</sup> and methodological features of pilot studies.<sup>7,14</sup> Seven evidence-based objectives of conducting such studies have been identified:

1. to evaluate the integrity of a study protocol for a larger study;
2. to acquire preliminary estimates for sample size computation;
3. to test data collection questionnaires;
4. to test the randomization technique(s);
5. to estimate the recruitment and consent rates;
6. to test the acceptability of the intervention; and
7. to choose the most suitable outcome measure(s).<sup>11</sup>

The extent to which pilot studies in the rehabilitation literature live up to these expectations is unknown and is the topic of this review using the pilot/feasibility studies published in *Clinical Rehabilitation* over the past three decades as examples. The specific objectives are to estimate the extent to which the identified pilot/feasibility studies: (i) address needed objectives; (ii) lead to definitive trials; and (iii) whether the subsequent undertaking of a definitive trial is influenced by the strength of the evidence of outcome improvement.

**Table 1.** Summary of previous reviews on the practices associated with pilot studies.

Study reference	Number and type of journal articles included	Time period of review
Lancaster et al. <sup>11</sup>	90; Four general clinical journals, <i>British Medical Journal</i> (BMJ), <i>Lancet</i> , <i>Journal of the American Medical Association</i> (JAMA), and <i>New England Journal of Medicine</i> (NEJM), and three subject-specific journals, the <i>British Journal of Cancer</i> (BJC), <i>British Journal of Obstetrics and Gynaecology</i> (BJOG) and the <i>British Journal of Surgery</i> (BJS)	2000–2001
Only four articles stated that the pilot study was in preparation for a clinical trial; >50% of the articles concluded that a further study was required.		
Araín et al. <sup>9</sup>	54; Follow-up to previous review; <sup>13</sup> describes same journals as given above, with the addition of 12 reports from the UK Clinical Research Network Portfolio Database	2007–2008
Of the studies, 48% were identified as ‘pilot’ and the rest were ‘feasibility’ studies; 81% focused on hypothesis testing and 81% highlighted the need for further study. Only eight of the 90 articles (9%) identified by the previous review <sup>11</sup> were followed up in a subsequent larger study.		
Shanyinde et al. <sup>12</sup>	50; Pilot and feasibility randomized controlled trials from EMBASE and MEDLINE databases	2000–2009
In only 56% of studies, (95% confidence intervals 41% to 70%) were methodological issues discussed in adequate details, 18% (95% confidence interval 9% to 30%) mentioned future trials in the discussion section, and only 12% (95% confidence interval 5% to 24%) of investigators were actually undertaking a subsequent trial.		
Kannan and Gowri <sup>13</sup>	93; Indian journals of allopathic medicine, dentistry, and complementary and alternative systems of medicine	Between January and December 2013
None of the studies presented the reason for piloting; none of them discussed feasibility; 2/3 of the articles did hypothesis testing and inferred the significance of differences between the groups and none of these studies mentioned power for these contrasts.		

## Methods

### Eligibility criteria

Trials published in the journal *Clinical Rehabilitation*, since its inception, were eligible if the word ‘pilot’ was specified anywhere in the article by the authors. Also eligible were research publications labelled as ‘feasibility studies’, ‘preliminary studies’, and ‘proof-of-concept studies’ as these terms are often used interchangeably with ‘pilot’ studies.<sup>9,14–16</sup> For the purpose of this review, the delineation between pilot and feasibility studies was not made, although the consensus definition<sup>8</sup> considers all preparatory studies as ‘feasibility’. The consensus reserves the term ‘pilot’ for those small-scale studies that have a specific design feature (either randomized or not) that test some or all aspects of a future trial. We use the term ‘pilot’ study to refer to the studies reviewed here and the term ‘feasibility’ if the authors of the chosen articles used this term rather than ‘pilot’.

### Search strategy

This study was embedded within a larger study reviewing the methodological features of trials published over the past 30 years in the journal *Clinical Rehabilitation*.<sup>3</sup> The search strategy has been described previously<sup>3</sup> and yielded 581 clinical trials that had a control group, randomized or not. Of these trials, 191 articles were identified to be pilot or feasibility studies by one of the 23 reviewers conducting the comprehensive data abstraction for the larger study.

### Data abstraction

For the review of full clinical trials,<sup>3</sup> a data abstraction form was devised and much of the data elements related to both full and pilot trials. A separate data extraction form was created specifically for pilot studies to include additional fields for: (i) location of declaration of the ‘pilot’ nature of the study; (ii) whether the authors stated the purpose of

the piloting; (iii) the reason(s) for the piloting; (iv) inferring the reason(s) if not clearly mentioned; (v) whether the study was followed up in a definitive clinical trial; (vi) whether the sample size calculation was made for the future definitive study; and (vii) data for calculating effect size for the between-group comparisons.

Both data abstraction forms included information on sample size at randomization and at the end of intervention and, hence, this information was abstracted by two reviewers. For the other fields not duplicated for full and pilot trials, the data extraction was conducted by one reviewer (NK), areas that lacked clarity were discussed with a senior reviewer (NM), and a decision was made as to the data to be abstracted. The information on the follow-up of a pilot study was obtained by examining its citations in SCOPUS. Additionally, the corresponding author of each article was sent an email inquiry to verify if a definitive study was undertaken. If an article did not provide an email address for correspondence, the follow-up status was decided based upon SCOPUS entries.

### Data analysis

Frequency distributions and means and standard deviations were used to describe the features of the pilot studies. Between-group effect size was computed for each pilot study using the standardized mean difference. Effect size was not calculated for studies that provided only median and interquartile range values. Based on the distribution of data, effect size was classified into six categories ( $\leq 0.1$ ;  $> 0.1$  and  $\leq 0.2$ ;  $> 0.2$  and  $\leq 0.5$ ;  $0.5$  and  $\leq 0.8$ ;  $> 0.8$  and  $\leq 2.0$ ; and  $> 2.0$ ). Logistic regression was used to estimate the association between the observed effect size and follow-up of pilot studies, with the effect size category  $\leq 0.1$  as the referent. Odds ratio (OR) and associated 95% confidence intervals (CI) were calculated. The pilot studies were divided into three eras based upon the year of publication ( $< 1999$ ;  $\geq 1999$  and  $\leq 2009$ ; and  $> 2009$  to  $\leq 2015$ ). Chi square analysis was employed to estimate the influence of era on the association between the strength of effect

and follow-up. All analyses were conducted in SAS 9.3.

## Results

A total of 191 pilot studies were identified for the time period 1987 to March 2015: seven (4%) before 1999; 71 (37%) between 1999 to 2009; and 113 (59%) after 2009 and up to the year 2015. Table 2 shows the key characteristics of the chosen studies. Pilot status was declared most often in the title (87%) and abstract (68%). The objective indicated pilot status in only 54% of studies. The purpose of piloting was specified in only 58% ( $n = 110$ ) of the pilot studies. Among these 110 studies, the most frequent purpose was to estimate the potential for efficacy (85%) followed by testing the feasibility of the intervention (60%). Feasibility of outcome measures, safety of intervention, sample size computation, and feasibility of recruitment rates were reported as one of the main objectives of piloting in  $< 11\%$  of studies. Other feasibility reasons that were identified included estimating retention rates<sup>17–20</sup> and testing the acceptability of intervention.<sup>20–23</sup>

For the 42% of studies ( $n = 81$ ) that did not declare a clear purpose for piloting, inference was based upon the information provided in the manuscripts. Estimation of efficacy potential (89%) and feasibility of intervention (15%) were the most common reasons followed by feasibility of recruitment rates and timing of intervention effect on outcomes.

Almost half of the studies used the terms ‘pilot’ and ‘feasibility’ interchangeably and only 6% of the 191 studies were uniquely labelled as ‘feasibility’ studies. Only 34% of the studies had done a power calculation for a future definitive clinical trial.

Table 3 indicates the sample sizes and drop-out proportions. The average sample size was 31 with a large range from 7 to 120. The proportion of drop-outs averaged 3% with a range from 0% to 31%.

Table 4 shows that, of the 191 pilot studies, 12% ( $n = 23$ ) were followed by a definitive clinical trial; an additional small percentage  $\sim 3.5\%$  ( $n = 7$ ) of studies had a main study underway or a published study protocol. The remaining (85%;

**Table 2.** Characteristics of the 191 pilot studies included in the review.

Characteristic	N	%
<i>Location of declaring 'pilot'</i>		
Title	168	87
Abstract	131	68
Title or abstract	184	96
Introduction/objective	104	54
Study design	78	41
Results	7	4
Discussion	105	55
Conclusion	49	26
<i>Purpose of 'piloting' declared</i>		
Yes	110	58
No	81	42
<i>Purpose of 'piloting' declared as<sup>a</sup></i>		
Feasibility of recruitment rates	9/110	8
Compliance or adherence rates	4/110	4
Timing of effect of an outcome	0/110	0
Feasibility of intervention	66/110	60
Feasibility of outcome measures	12/110	11
Estimation of efficacy potential	94/110	85
Safety of intervention	11/110	10
Computation of sample size	10/110	9
<i>Purpose inferred when not declared as<sup>a</sup></i>		
Feasibility of recruitment rates	5/81	6
Compliance or adherence rates	1/81	1
Timing of effect of an outcome	5/81	6
Feasibility of intervention	12/81	15
Feasibility of outcome measures	0/81	0
Estimation of efficacy potential	72/81	89
Safety of intervention	2/81	2
Computation of sample size	1/81	1
<i>Studies labelled as 'feasibility'</i>	12	6
<i>Power calculation made for a definitive clinical trial</i>	64	34

<sup>a</sup>Studies could have more than one reason for 'piloting'.

**Table 3.** Distribution of sample size in the included 191 studies.

	Total sample size	Drop-out proportions
Mean (SD)	31 (18.3)	3 (4.6)
25 percentile	20	0
50 percentile	28	2
75 percentile	40	5
Range	7–120	0–31

**Table 4.** Follow-up status of the 191 pilot studies.

	N	%
<i>Follow-up of pilot studies</i>		
Completed	23	12
Trial underway or completed but not yet published	4	2
Published protocol available	3	1.5
None	162	85
Email contact available	173	90
<i>Outcome of email contact</i>		
Undelivered	76/173	44
Unanswered	26/173	15
Answered	71/173	41
<i>Reasons presented for no follow-up</i>		
Lack of funding	9/17	53
Results confirmed by another team	2/17	12
Pilot work conducted as a part of student's thesis	2/17	12
Principal investigator no longer doing research	2/17	12
Product to be evaluated not made available	1/17	6
Power analysis indicated recruitment not feasible	1/17	6

$n = 162$ ) did not appear to have had any further follow-up. There was no effect of era on follow-up status (data not shown). Also demonstrated in Table 4 is the follow-up process. A total of 173 emails were sent, enquiring about follow-up status. Of these, 44% ( $n = 76$ ) were not delivered, 15% ( $n = 26$ ) were unanswered, and the remaining 41% ( $n = 71$ ) were answered. Only 17 of the corresponding authors provided a reason for non-pursuance of their pilot work. The commonly encountered cause for no follow-up was the lack of a funding resource.

Table 5 presents the association between effect size and follow-up of the 144 studies that presented sufficient data to allow computation of between-group effect size. Logistic regression analyses demonstrated that in comparison with the lowest effect size, i.e. Cohen's  $d$  less than 0.1, there was a tendency for a lower odds of follow-up for studies with effect sizes between 0.8 and to 2.0 to be followed up (OR 0.69, CI 0.08 to 5.46). Studies with effect sizes greater than 2.0 had the same odds of

**Table 5.** Association between effect size and follow-up of 144 studies.<sup>a</sup>

Effect size (Cohen's d)	Total number of studies	Number of studies with follow-up	Odds ratio	Confidence intervals
	N (%)	N (%)		
No data	47 (25)	8 (17)		
<0.1	18 (13)	2 (11)	Referent	
>0.1 to ≤0.2	21 (15)	6 (29)	3.20	0.05 to 18.38
>0.2 to ≤0.5	33 (23)	5 (15)	1.42	0.24 to 8.23
>0.5 to ≤0.8	28 (19)	5 (18)	1.73	0.29 to 10.1
>0.8 to 2	25 (17)	2 (8)	0.69	0.08 to 5.46
>2	19 (13)	2 (11)	0.94	0.11 to 7.49

<sup>a</sup>Studies with sufficient data to estimate between-group effect size.

follow-up as those with very small effect sizes, <0.1 (OR 0.94, CI 0.11 to 7.49). However, none of the associations were statistically significant as the CIs included the null value of 1.0.

## Discussion

This article reviewed the state of pilot/feasibility studies published in *Clinical Rehabilitation* since its inception (1987). During the 30-year time period, 191 pilot studies were published, while the corresponding number of full trials was 390, indicating the importance of pilot trials in the rehabilitation literature. The specific purposes of piloting were not always stated. In fact, only 58% ( $n = 110$ ) of the studies clearly declared what was being piloted. The most frequent purpose was to estimate the potential for efficacy (85%) followed by testing the feasibility of the intervention (60%). The terms 'pilot' and 'feasibility' were often used interchangeably to describe studies designed to inform future trials. However, only 12% of these studies were followed by a definitive clinical trial and <4% of studies had a main study underway or a published study protocol.

This review identified, in one journal only and over 30 years, that one-third of clinical trials were pilot trials. An Ovid MEDLINE search of the term 'pilot study' in the titles of the research articles published over only 1 year (2014) yielded 6002 records, indicating the widespread use of the term. However, the previous corpus of

reviews on the state of pilot studies (see Table 1) has indicated that not enough justice has been done to their conduct or follow-up.

### *What should be the focus for the next 30 years of pilot studies?*

**Label pilot studies correctly.** The recently devised conceptual framework for the definitions of preparatory studies<sup>8</sup> and the CONSORT reporting guidelines for pilot and feasibility studies<sup>24</sup> should be followed. As proposed by the conceptual framework, when there is uncertainty about the feasibility of a future randomized controlled trial, a 'feasibility' study should be carried out.<sup>8</sup> Not addressed by this framework<sup>8</sup> are internal pilot studies, which are fundamentally part of the definitive trial, but are mainly used to revise the sample size estimates upwards based on initial effect size estimates.<sup>11</sup> In rehabilitation, these internal pilot studies would be rare.

As per the framework, 'feasibility' is an umbrella term that encompasses three types of preparatory studies.<sup>8</sup> Randomized pilot studies are those in which a future definitive clinical trial involving randomized study groups or its components are investigated on a miniature scale. Non-randomized pilot studies are similar to the randomized pilot studies except these do not include randomization of study participants. Another category includes feasibility studies that are not pilot studies. These endeavour to test

whether some component of a future trial can be executed and may address the development of an intervention in some manner. However, these do not involve implementing an intervention or other components associated with processes that may be needed to be carried out in a future main study.<sup>8</sup>

This review found only 6% of the included studies were labelled as 'feasibility'. Mostly, the terms 'pilot' and 'feasibility' were employed without any distinction between the two. Consistency in the terms used to label the preparatory studies should improve with the advent of the consensus definitions<sup>8</sup> and the CONSORT reporting guidelines for pilot and feasibility studies.<sup>24</sup> Also, declaration of pilot/feasibility status both in the title or abstract is deemed useful for indexing purposes and for an easy identification in the electronic database searches.<sup>7,8</sup> This is recommended by the CONSORT guidelines<sup>24</sup> as well. Most of the reviewed pilot studies complied with this recommendation, whereas a minor proportion designated the status in the conclusion section only. Better labelling of the pilot studies can improve their impact and visibility.

*Distinguish pilot studies from small clinical trials.* It has been recognized that authors designate small clinical trials as 'pilot' studies 'post-hoc' when it is clear that they cannot reach a definitive answer from the data accrued; some reviewers or journal editors also insist on this labelling, although the trial was not developed as 'pilot'.<sup>12</sup> Consequently, small trials primarily estimating efficacy end up being labelled as 'pilot' without the objectives compatible with pilot/feasibility status. This practice has also been adopted by *Clinical Rehabilitation*. As suggested by Sackett and Cook over two decades ago,<sup>25</sup> methodologically sound small clinical trials can lead to vital lessons. They have a potential to challenge traditional therapeutic judgements that have not been put to investigation before. Therefore, they should be labelled as such and not disguised as 'pilot'. Many prominent journals, including *Clinical Rehabilitation*, do not yet have an existing policy for the conduct and reporting of pilot studies.<sup>9</sup> Changing journal policy about reporting of pilot studies would help improve the

situation. Recently, a new open access journal called *Pilot and Feasibility Studies* has been created to ensure that all foundational work conducted for large-scale studies can be brought to light.<sup>26</sup>

*Focus on the required objectives of pilot/feasibility studies.* Based on the pilot studies published in *Clinical Rehabilitation*, it was found that most were undertaken with specific objectives of estimating the potential for efficacy and testing the feasibility of intervention. Indicators of feasibility did not receive much attention in the reviewed studies. This is in line with the previous reviews.<sup>12</sup> One aspect of feasibility that is unique to rehabilitation studies is the feasibility of the measurement strategy. Unlike the pharmacological trials where the outcomes are mostly directly measured, rehabilitation studies tend to have multiple outcomes that are a mix of directly measured as well as patient-reported, and can also be measures of complex, theoretical constructs, such as health-related quality of life.<sup>4</sup> Therefore, it is crucial to test the feasibility of the outcome measurement strategy to avoid missing data arising from a measurement approach that is too burdensome for respondents. As outlined by one of the most comprehensive guides on pilot studies,<sup>7</sup> objectives should address process (e.g. recruitment, refusal, retention, and adherence rates), resources (e.g. adequacy of equipment), and management (e.g. data handling) issues. Our review found scant emphasis on these objectives (see Table 2). Including all this information in the pilot phase not only facilitates the conduct of a full-strength trial, but also leads to a more competitive proposal for funding purposes.<sup>10,27</sup>

*Justify the rationale for chosen sample size.* In terms of sample size, there was an average of 31 participants in the pilot studies reviewed in this manuscript, with a wide range from seven to 120 participants. A group of researchers conducted an audit of sample sizes in pilot studies carried out in the UK and reported a comparable range, from eight to 114 participants. On the other hand, feasibility trials had a minimum of 10 to a maximum of 300 participants.<sup>28</sup> Several authors have made recommendations for sample size to be included. For

example, there is a general rule of thumb to recruit at least 30 participants or higher for parameter estimation,<sup>29</sup> whereas another researcher<sup>30</sup> suggests recruiting at least 12 participants per study group. To minimize the imprecision surrounding the estimation of standard deviation, a total sample size of 70 participants is deemed necessary in a study with two treatment arms.<sup>31</sup> Although rationale for establishing sample size is important to be included, there is a view that a formal calculation may not be appropriate.<sup>28</sup> For example, if the intention is to support feasibility of the main trial from adherence or completion rates from a pilot, the confidence one can have from these pilot estimates is a function of the pilot sample size. If a pilot study of 30 people has observed a completion rate of 80%, the 95% confidence around this proportion is 63% to 90%. This means that a definitive trial is more likely to have a lower completion rate than higher; even increasing the pilot size to 50 participants would not yield much greater estimation confidence (95% CI 67% to 90%).<sup>32</sup>

*Use the correct analytic approach.* It is essential to underscore that pilot studies are not intended for testing hypothesis,<sup>16</sup> however, they can certainly indicate the potential for efficacy,<sup>24</sup> which would support pursuing a definitive trial. Between-group comparisons should not be performed, as the study, by design, is not powered for this contrast.<sup>16</sup> Ideally, the authors should report descriptive statistics, point estimates, and CI for the effect observed.<sup>11</sup> As recommended by Lee and his colleagues,<sup>33</sup> CI should be interpreted in relation to a priori-determined minimally important differences (MID).<sup>16,34</sup>

Effect size is the parameter that best indicates the potential for efficacy.<sup>24</sup> Pilot studies with effect sizes in the small or trivial range could be considered to provide weak evidence of efficacy potential. It is vital to mention that only 144 studies could be included in the main analysis, as the rest of the studies did not provide data for the computation of the effect size. Most of such studies presented either median and interquartile range values or had included no data that could lead to the effect size estimation. In the future, researchers need to ensure that ample data are given in the manuscript.

Approaches to statistical analysis that go beyond simply reporting mean changes, such as defining responder-status, are also recommended.<sup>3</sup> This particular approach involves dichotomizing a continuous primary outcome measure into ‘responders’ and ‘non-responders’ based on a magnitude of effect deemed to be important. This information can be used to enhance interpretability of the data collected and can provide preliminary estimates of number needed to treat.<sup>10</sup>

*Improve on the reporting of pilot studies.* The recently generated reporting guidelines for pilot and feasibility studies, produced as an extension for the CONSORT statement,<sup>24</sup> should be implemented, however, additional emphasis is likely needed for rehabilitation studies. Most rehabilitation interventions require pilot testing as these interventions tend to be complex owing to the need for tailoring to the individual, their multi-modal nature, the number of other active ingredients,<sup>4</sup> such as people, setting, and attention, that need to be balanced by the control situation,<sup>3</sup> and the difficulty in masking research personnel and participants<sup>2</sup> when assessing outcomes. In these complex situations, pilot studies serve a crucial role of ensuring a robust methodological approach in a subsequent definitive trial. There is some guidance available on how to approach pilot and feasibility studies associated with occupational therapy interventions systematically.<sup>4</sup> To improve the reporting standards in rehabilitation in general, the authors of pilot studies should clearly incorporate information on the success of randomization, suitability of control condition, optimal recruitment strategy, drop-out rates, intervention integrity (i.e. if the intervention is delivered as per the original plan to each participant),<sup>35</sup> adverse events, and power calculation for a full-strength trial, in addition to the other aspects of feasibility. Appropriate objectives should be stated explicitly. The authors should acknowledge that the validity of findings could be dubious if they employ inferential statistics in a pilot study with a small sample size.<sup>16</sup> And for future studies, the data collected from pilots serve two purposes: Inform the need for a future trial and the determination of sample size required to confirm the hypothesis.



*Justify the need for a further trial, and do it.* One of the main objectives of this particular review was to estimate the extent to which the pilot study outcomes influenced the undertaking of a subsequent definitive trial. Only a small proportion of pilot studies were eventually followed up and these findings concur with previous reports.<sup>9</sup> Contrary to the usual expectations, it was found that the strength of effect observed in pilot studies was not associated with the follow-up status. The thinking here was that if the effect was nil or close to nil, a main trial may not be justified as there was no evidence for efficacy potential. On the other hand, pilot studies with very large effect sizes may also not progress to the definitive stage as the authors and funders may think it is no longer ethical to offer the control condition. Cronbach counselled against thinking that the results of pilot studies will be replicated in a larger study by introducing 'superrealization bias'. It is an outcome of the observation that in small scale studies, the researchers are capable to attain a high quality implementation which could never be achieved on a larger scale.<sup>36</sup> Only one pilot study was identified as addressing the issue indicating that the effect size found in the small pilot on a walking intervention for cancer fatigue was 'over optimistic'.<sup>37</sup>

Funding agencies often require that a pilot study is truly encouraging to persuade funders that a full strength clinical trial should be carried out.<sup>10,38</sup> Although the reasons for not being able to conduct a subsequent trial were not probed in this review, two of the corresponding authors pointed out that lack of funding hindered them from pursuing their work. Among other reasons, was change-over in personnel, particularly if the pilot was undertaken by a trainee.

*Be vigilant while calculating sample size needed in next trial.* Over 60% of reviewed studies did not provide sample size estimates for a subsequent full-scale trial using the data accrued from the pilot. Although sample size computation for a future trial is considered one of the fundamental objectives of pilot studies, it is worth noting that small datasets tend to yield imprecise effect size estimates, so the sample size estimates should

be interpreted with caution.<sup>28,38,39</sup> In an observational study by Salbach and her colleagues,<sup>40</sup> the responsiveness of gait speed over time had an effect size of 1.22 with CIs (0.93 to 1.50). With a total of 50 participants, the width of the CI is quite wide and choosing the mid-point rather than the lower bound would greatly affect the sample size projections. Sample size estimates should not be based solely on the size needed to reject the null hypothesis (no effect), rather the size needed to reject a trivial alternative. This requires estimating sample size for a desired CI, one that excludes an effect of smaller than say 0.2 of a standard deviation (Cohen's effect size of 0.2).

### Limitations

The studies incorporated in this review were acquired from a companion review of clinical trials.<sup>3</sup> An independent, systematic search of pilot/feasibility studies published in the journal was carried out subsequently and identified no additional studies. There is a likelihood of reviewer bias as not all elements of data abstraction were validated by a second reviewer. Also, it is possible that the information on the follow-up of pilot studies may not have been captured correctly, as citation entries may not be updated in SCOPUS. However, email enquiries were sent to the corresponding authors for the information on follow-up. Moreover, the studies that had no citations in SCOPUS were also searched in PubMed.

Although feasibility and pilot studies build a rich groundwork for definitive clinical trials, the practise surrounding their conduct and reporting needs to be bettered. For clarity and uniformity in the definitional context, the recently devised framework<sup>8</sup> for preparatory studies should be implemented. The notorious tradition of presenting the small clinical trials as pilot studies on account of their meagre sample size or other flaws should be avoided. Researchers need to ensure that methodologically sound pilot studies are undertaken and multiple aspects of workability of study protocol are investigated so that potential bumps in the road can be dealt with before embarking upon a full strength clinical trial.

## Key messages

As there is likely considerable uncertainty when planning a full-scale trial of a rehabilitation intervention, feasibility needs to be tested and demonstrated prior to committing considerable human and monetary resources.

Feasibility is the overarching term encompassing: (i) pilot trials (randomized or not) testing both the intervention and other aspects of the trial process; and (ii) other feasibility studies that mainly test a process or may address development of an intervention.

The recently issued reporting guidelines for pilot and feasibility studies need to be followed by researchers, reviewers, and journals alike.

Small studies or studies that go wrong should not be labelled ‘pilot’ after the fact.

Pilot studies should fully describe the distribution of the sample on all outcomes at all time points, and provide point estimates of change with CIs rather than only a *p*-value.

Effect sizes estimated from pilot studies should be interpreted as potentially over-optimistic and power calculations should be adjusted accordingly.

## Conflict of interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

## References

1. Mayo NE. *Randomized trials and other parallel comparisons of treatment*. John Wiley & Sons, Inc & The New England Journal of Medicine, 2009, pp.51–89.
2. Andrews K. The limitations of randomized controlled trials in rehabilitation research. *Clin Rehabil* 1991; 5(1): 5–8.
3. Mayo NE, Kaur N, Barbic SP, et al. How have research questions and methods used in clinical trials published in *Clinical Rehabilitation* changed over the last 30 years? *Clin Rehabil* 2016; 30(9): 847–864.
4. Tickle-Degnen L. Nuts and bolts of conducting feasibility studies. *Am J Occup Ther* 2013; 67(2): 171–176.
5. National Institute for Health Research. *Feasibility and Pilot Studies*. 2016. [http://www.nihr.ac.uk/CCF/RfPB/FAQs/Feasibility\\_and\\_pilot\\_studies.pdf](http://www.nihr.ac.uk/CCF/RfPB/FAQs/Feasibility_and_pilot_studies.pdf) (accessed on June 2016).
6. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I and Petticrew M. Developing and evaluating complex interventions: The new Medical Research Council guidance. *Int J Nurs Stud* 2013; 50(5): 587–592.
7. Thabane L, Ma J, Chu R, et al. A tutorial on pilot studies: The what, why and how. *BMC Med Res Methodol* 2010; 10: 1.
8. Eldridge SM, Lancaster GA, Campbell MJ, et al. Defining feasibility and pilot studies in preparation for randomised controlled trials: Development of a conceptual framework. *PLoS One* 2016; 11(3): e0150205.
9. Arain M, Campbell MJ, Cooper CL and Lancaster GA. What is a pilot or feasibility study? A review of current practice and editorial policy. *BMC Med Res Methodol* 2010; 10: 67.
10. Moore CG, Carter RE, Nietert PJ and Stewart PW. Recommendations for planning pilot studies in clinical and translational research. *Clin Transl Sci* 2011; 4(5): 332–337.
11. Lancaster GA, Dodd S and Williamson PR. Design and analysis of pilot studies: Recommendations for good practice. *J Eval Clin Pract* 2004; 10(2): 307–312.
12. Shanyinde M, Pickering RM and Weatherall M. Questions asked and answered in pilot and feasibility randomized controlled trials. *BMC Med Res Methodol* 2011; 11: 117.
13. Kannan S and Gowri S. Pilot studies: Are they appropriately reported? *Perspect Clin Res* 2015; 6(4): 207–210.
14. Arnold DM, Burns KE, Adhikari NK, Kho ME, Meade MO and Cook DJ. The design and interpretation of pilot trials in clinical research in critical care. *Crit Care Med* 2009; 37(1 Suppl): S69–S74.
15. Morin K. Value of a pilot study. *J Nurs Educ* 2013; 52(10): 547–548.
16. Leon AC, Davis LL and Kraemer HC. The role and interpretation of pilot studies in clinical research. *J Psychiatr Res* 2011; 45(5): 626–629.
17. Cameron JI, Naglie G, Green TL, et al. A feasibility and pilot randomized controlled trial of the “Timing it Right Stroke Family Support Program”. *Clin Rehabil* 2014; 29(11): 1129–1140.
18. Morgan P, Murphy A, Opheim A, Pogrebnov D, Kravtsov S and McGinley J. The safety and feasibility of an intervention to improve balance dysfunction in ambulant adults with cerebral palsy: A pilot randomized controlled trial. *Clin Rehabil* 2014; 29(9): 907–919.
19. Crosbie JH, Lennon S, McGoldrick MC, McNeill MD and McDonough SM. Virtual reality in the rehabilitation of the arm after hemiplegic stroke: A randomized controlled pilot study. *Clin Rehabil* 2012; 26(9): 798–806.
20. Minns Lowe CJ, Barker KL, Holder R and Sackley CM. Comparison of postdischarge physiotherapy versus usual care following primary total knee arthroplasty for

- osteoarthritis: An exploratory pilot randomized clinical trial. *Clin Rehabil* 2012; 26(7): 629–641.
21. Jepsen P, Sands G, Beswick AD, Davis ET, Blom AW and Sackley CM. A feasibility randomised controlled trial of pre-operative occupational therapy to optimise recovery for patients undergoing primary total hip replacement for osteoarthritis (PROOF-THR). *Clin Rehabil* 2015; 30(2): 156–166.
  22. Elsworth C, Winward C, Sackley C, et al. Supported community exercise in people with long-term neurological conditions: A phase II randomized controlled trial. *Clin Rehabil* 2011; 25(7): 588–598.
  23. Kloos AD, Fritz NE, Kostyk SK, Young GS and Kegelmeyer DA. Video game play (Dance Dance Revolution) as a potential exercise therapy in Huntington's disease: A controlled clinical trial. *Clin Rehabil* 2013; 27(11): 972–982.
  24. Eldridge SM, Chan CL, Campbell MJ, et al. CONSORT 2010 statement: Extension to randomised pilot and feasibility trials. *Pilot Feasibility Stud* 2016; 2: 64.
  25. Sackett DL and Cook DJ. Can we learn anything from small trials? *Ann NY Acad Sci* 1993; 703(1): 25–32.
  26. Lancaster GA. Pilot and feasibility studies come of age! *Pilot Feasibility Stud* 2016; (1): 1.
  27. Lancaster GA, Campbell MJ, Eldridge S, et al. Trials in primary care: Statistical issues in the design, conduct and evaluation of complex interventions. *Stat Methods Med Res* 2010; 19(4): 349–377.
  28. Billingham SA, Whitehead AL and Julious SA. An audit of sample sizes for pilot and feasibility trials being undertaken in the United Kingdom registered in the United Kingdom Clinical Research Network database. *BMC Med Res Methodol* 2013; 13: 104.
  29. Browne RH. On the use of a pilot sample for sample size determination. *Stat Med* 1995; 14: 1933–1940.
  30. Julious SA. Sample size of 12 per group rule of thumb for a pilot study. *Pharma Stats* 2005; 4(4): 287–291.
  31. Teare MD, Dimairo M, Shephard N, Hayman A, Whitehead A and Walters SJ. Sample size requirements to estimate key design parameters from external pilot randomised controlled trials: A simulation study. *Trials* 2014; 15: 264.
  32. Newcombe RG. Two-sided confidence intervals for the single proportion: Comparison of seven methods. *Stat Med* 1998; 17(8): 857–872.
  33. Lee EC, Whitehead AL, Jacques RM and Julious SA. The statistical interpretation of pilot trials: Should significance thresholds be reconsidered? *BMC Med Res Methodol* 2014; 14: 41.
  34. Mayo NE (ed.). *Dictionary of quality of life and health outcomes measurement*. Milwaukee, WI: International Society for Quality of Life Research, 2015.
  35. Conn VS, Algase DL, Rawl SM, Zerwic JJ and Wyman JF. Publishing pilot intervention work. *West J Nurs Res* 2010; 32(8): 994–1010.
  36. Cronbach LJ, Ambron SR, Dornbusch SM, et al. *Toward reform of program evaluation: Aims, methods, and institutional arrangements*. San Francisco, Jossey-Bass, 1980.
  37. Mayo NE, Moriello C, Scott SC, Dawes D, Auais M and Chasen M. Pedometer-facilitated walking intervention shows promising effectiveness for reducing cancer fatigue: A pilot randomized trial. *Clin Rehabil* 2014; 28(12): 1198–1209.
  38. Friedman L. Commentary: Why we should report results from clinical trial pilot studies. *Trials* 2013; 14: 14.
  39. Sim J and Lewis M. The size of a pilot study for a clinical trial should be calculated in relation to considerations of precision and efficiency. *J Clin Epidemiol* 2012; 65(3): 301–308.
  40. Salbach NM, Mayo NE, Higgins J, Ahmed S, Finch LE and Richards CL. Responsiveness and predictability of gait speed and other disability measures in acute stroke. *Arch Phys Med Rehabil* 2001; 82(9): 1204–1212.