



i-CONTENT tool for assessing therapeutic quality of exercise programs employed in randomised clinical trials

Thomas J Hoogeboom ¹, Martijn C Kousemaker,² Nico LU van Meeteren,³ Tracey Howe,⁴ Kari Bo ^{5,6}, Peter Tugwell ⁷, Manuela Ferreira,⁸ Rob A de Bie,⁹ Cornelia HM van den Ende,¹⁰ Jennifer E Stevens-Lapsley^{11,12}

For numbered affiliations see end of article.

Correspondence to

Dr Thomas J Hoogeboom, Radboud Institute for Health Sciences, IQ healthcare, Radboudumc, Nijmegen, Gelderland, Netherlands; thomas.hoogeboom@radboudumc.nl

Accepted 6 October 2020

ABSTRACT

Objective When appraising the quality of randomised clinical trial (RCTs) on the merits of exercise therapy, we typically limit our assessment to the quality of the methods. However, heterogeneity across studies can also be caused by differences in the quality of the exercise interventions (ie, 'the potential effectiveness of a specific intervention given the potential target group of patients')—a challenging concept to assess. We propose an internationally developed, consensus-based tool that aims to assess the quality of exercise therapy programmes studied in RCTs: the international Consensus on Therapeutic Exercise and Training (i-CONTENT) tool.

Methods Forty-nine experts (from 12 different countries) in the field of physical and exercise therapy participated in a four-stage Delphi approach to develop the i-CONTENT tool: (1) item generation (Delphi round 1), (2) item selection (Delphi rounds 2 and 3), (3) item specification (focus group discussion) and (4) tool development and refinement (working group discussion and piloting).

Results Out of the 61 items generated in the first Delphi round, consensus was reached on 17 items, resulting in seven final items that form the i-CONTENT tool: (1) patient selection; (2) qualified supervisor; (3) type and timing of outcome assessment; (4) dosage parameters (frequency, intensity, time); (5) type of exercise; (6) safety of the exercise programme and (7) adherence to the exercise programme.

Conclusion The i-CONTENT-tool is a step towards transparent assessment of the quality of exercise therapy programmes studied in RCTs, and ultimately, towards the development of future, higher quality, exercise interventions.

therapy interventions may differ in modes, dosage and administration, all of which will impact their quality, and consequently, their therapeutic potential. One might argue that it is almost unethical that researchers are still able, without any regulation, to design and test exercise interventions that likely have a low potential for effectiveness. There is an urgent need for an explicit tool that will assess the quality of an exercise intervention.^{7–11} The international Consensus on Therapeutic Exercise and Training (i-CONTENT) tool for assessing the quality of exercise interventions aims to make this possible.

Challenges and shortcomings in exercise therapy evaluations

Exercise therapy is used by patients with support or supervision from physiotherapists, exercise scientists and rehabilitation physicians. In the scientific field of exercise therapy, interventions are often poorly described.^{10 12–17} Over the last decade, a number of reporting guidelines have been published in the field of exercise therapy with the intent to improve the reproducibility of exercise interventions in scientific papers, like the Consolidated Standards of Reporting Trials statement,¹⁸ the Standard Protocol Items: Recommendations for Interventional Trials statement,¹⁹ the Template for Intervention Description and Replication (TIDieR) checklist,²⁰ and the Consensus on Exercise Reporting Template (CERT).²¹

Although adequate reporting of exercise interventions is assumed to be crucial to the understanding and reproducibility of interventions, it still does not help the reader to determine the quality (ie, 'the potential effectiveness of a specific intervention given the potential target group of patients') of an exercise intervention. Moreover, it does not help the end-users—patients, professionals, health-care financiers to weigh, choose and appreciate the different intervention options. A well-documented exercise intervention can still be of low therapeutic quality. In an earlier attempt to evaluate the quality of exercise programmes using a locally developed tool⁹ we found that of 57 assessed trials (comprising over 4500 volunteers), 88% evaluated suboptimal exercise programmes, which were unlikely to yield meaningful clinical results.²² Assessing the quality of exercise interventions is one of the major challenges in the field of exercise therapy research.^{7–11}

INTRODUCTION

Most people who are at risk of no longer being able to self-manage¹ can benefit from therapeutic exercise,^{2–4} under the prerequisite that the exercise programme is of sufficient quality.⁵ The scientific exercise community has an obligation when applying and advancing scientific knowledge, to maximise direct and indirect benefits to patients, research participants and other affected individuals, while minimising harm.⁶ However, in 2005, Herbert and Bø argued that not every exercise intervention tested in a randomised clinical trial (RCT) is of similar quality.⁷ After all, exercise



© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Hoogeboom TJ, Kousemaker MC, van Meeteren NLU, et al. *Br J Sports Med* Epub ahead of print: [please include Day Month Year]. doi:10.1136/bjsports-2019-101630

The currently available reporting tools do not interpret the quality of exercise interventions.¹⁰

Aim and scope

The aim of the i-CONTENT working group was to provide recommendations, in the form of a single useful rating and appraisal tool, to rate the quality of exercise therapy interventions, while taking previous efforts into account.^{9 18–21} The pool of potential users of the i-CONTENT tool are researchers developing, reporting or reviewing exercise therapy evaluations, and editors and peer reviewers evaluating publications on exercise therapy, while the wider audience might be patients, healthcare professionals and financiers working with exercise therapy. We believe the tool (which consists of a 7-item checklist) is a useful and practical tool for these initiators and audiences and will improve our understanding of the quality of exercise interventions and, ultimately, our individual and collective thoughts about attributions and contributions of these interventions to exercise therapy outcomes.

METHODS

The tool was developed by the i-CONTENT working group. The eight i-CONTENT working group members were purposefully sampled by the primary author (TJH) based on their long-standing academic expertise and contribution to the field of exercise therapy research. All members of the working group had a PhD: seven members were specialised in sports medicine, exercise therapy or physiotherapy practice (NLvM, RdB, TH, CHvdE, JES-L, MF and KB) and two in clinical epidemiology (PT and RAdB). Six members were active in the Cochrane Collaboration (MF, PT, TH, KB, NLvM and RdB). Finally, four members served as editors for journals in related fields (PT, JES-L, MF, RdB).

The i-CONTENT working group followed a four-stage Delphi approach to develop the i-CONTENT tool: (1) item generation (Delphi round 1), (2) item selection (Delphi round 2 and 3), (3) item specification (focus group discussion) and (4) tool development and refinement (working group discussion and piloting) (see figure 1). The working definition of therapeutic quality was ‘the potential effectiveness of a specific intervention given the potential target group of patients’.⁹ Exercise therapy was defined as ‘a regimen or plan of physical activities designed and prescribed for specific therapeutic goal’ (Mesh database). The results from the four stages were compiled to create the tool.

Stage 1: generating an item pool

To ensure 30 responders in the last round, previous Delphi studies suggest that in a worst case scenario, 80 responders would be needed to participate in the first Delphi round²³ and in a best case scenario, 43 responders.²⁴ It was expected that

60 responders for the first Delphi questionnaire would suffice to include at least 30 responders in the last round. We included experts in the field of physiotherapy, exercise therapy, exercise physiology, clinical medicine and clinical research, allowing for a heterogeneous group of experts.²⁵ The initial selection of experts was done after a pragmatic PubMed search; search terms “randomized clinical trials”, “exercise”, and “JAMA, BMJ, NEJM, Lancet, or PTJ” with the following limits were used: Adults (age >18 years) and publication year >2009. The first author of papers that studied the effectiveness of therapeutic exercise (exercise had to be the main intervention) in an RCT were contacted. Consequently, we asked these experts who they, outside their own research group, considered experts in the area of therapeutic exercise.²⁶ The aim was to include ‘in depth-experts’,²⁷ from a group selected on their work and achievements rather than acquaintances,²⁸ and provoke a snowball effect to efficiently include the 60 responders. Experts were invited by email to participate in the study. Anonymity among experts was maintained throughout all Delphi rounds.

In the first round, we asked questions about the participants’ demographics (ie, age, sex, education and profession), participants’ level of expertise (ie, regarding scientific output on therapeutic exercise) and therapeutic quality. Questions related to therapeutic quality asked during the first Delphi round are shown in table 1. Data saturation was assessed by checking whether new surveys revealed new items.²⁹

Stage 2: item selection

For the second round, the first author and a PhD student (JES-L) collated and grouped the responses from round one into a number of statements regarding therapeutic quality in exercise therapy. Consequently, the Delphi group was asked which of the statements they deemed essential for this rating scale (one point=very unessential, through to seven points=very essential).

In the third Delphi round, personalised questionnaires were created by the second author for each of the experts. These questionnaires comprised the median and iIQR of scores of each statement (representing group level of agreement and the degree of consensus, respectively) and their own personal rating. All experts reviewed and rerated all statements.

Finally, the second author prepared a list of statements which achieved consensus. Consensus for inclusion was defined a priori as a median rating of 6 or 7 on the 7-point rating scale and an IQR of 1.5 or less.³⁰

Stage 3: item specification

After applying the cut-off values to the items from the third Delphi round, a focus group was held to prepare a survey for the i-CONTENT working group to collate the remaining items into a tool. This focus group discussed the following topics: (1)

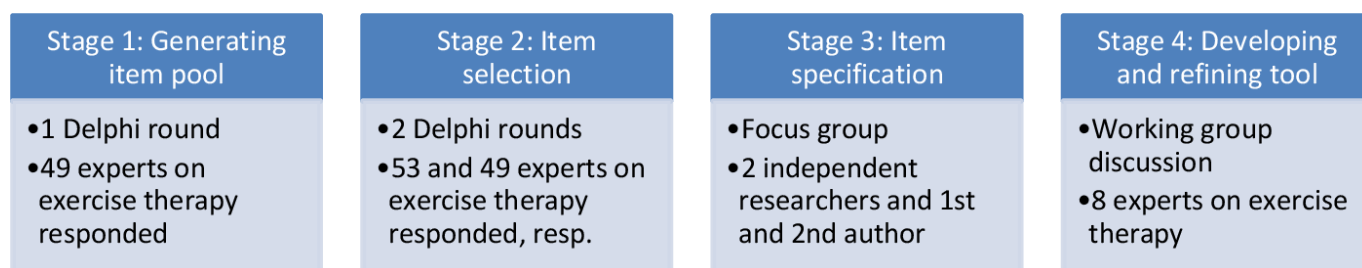


Figure 1 The four stages of the consensus procedure to develop the i-CONTENT tool. i-CONTENT, international Consensus on Therapeutic Exercise and Training.

Table 1 Questions related to the therapeutic quality in the first Delphi round

Questions from the first delphi round	
1	Should (randomised) trials on therapeutic exercise be criticised on their therapeutic validity?
2	Should the appraisal of therapeutic validity be included in a systematic review/meta-analysis on exercise therapies?
3	What characterises a therapeutically valid exercise therapy in your opinion?
4	What do you consider critical success criteria for therapeutic validity in exercise therapy?
5	What should be reported in a scientific paper to be able to address the therapeutic validity of the exercise therapy?
6	In what form do you think therapeutic validity of an exercise therapy should be scored/assessed? Checklist, rating scale or reporting tool.
7	If you have any comments/concerns/questions regarding this questionnaire, please let us know

are there similar items, (2) are there items which can be covered by a similar item and (3) are there items which are multi-interpretatable. The focus group comprised two independent researchers from the Radboudumc (BS and RN) and the first and second author. These researchers were selected using purposive sampling. None of the researchers were included in the Delphi study and were all educated on the subject of exercise therapy. The entire discussion was recorded using a Roland R-05 hand-held audio recorder.³¹ The second author transcribed the discussion to extract the conclusions. The two researchers were asked to give their opinion and agree to the conclusions extracted from the recordings and the transcription.

Survey

Following the focus group meeting, the first and second author created a survey for the working group. The survey contained a categorisation of the items, the conclusions from the discussion, and the question to submit two papers on exercise therapy. To make sure the participants were familiar with the items, the survey started off with the question to categorise the items in a way they deemed fit. The survey was sent to the i-CONTENT working group. Proposed changes were implemented when at least 75% (at least 6 out of 8) of the group members agreed.³²

Stage 4: developing and refining of the tool

A working group discussion was planned to discuss the outcome of the survey, as well as to come to a prefinal concept of the i-CONTENT tool. Prior to the discussion, the participants received a document containing the previous developments, the original items from the Delphi rounds, a concept for the tool, and outstanding discussion points from the survey. The results from the discussion were summarised and sent to all participants to receive their input, as not everyone would be able to participate due to time zone differences. Additional results were obtained via email. Consensus was reached if at least six out of the eight group members agreed to the proposed changes.

Finally, to test the tool's interpretability, the second author and a PhD student from Caledonian University (JG) piloted the prefinal version of the tool. Seven articles on exercise therapy for people with shoulder complaints were selected at random from a larger systematic review that is in preparation. The second author and JG independently scored the articles and discussed in an online meeting their experiences using the checklist. Results

from the discussion were used to refine the checklist to its final state.

RESULTS

During the first Delphi round, 65 people were initially invited. Participants were asked to suggest others to participate, which led to the invitation of another 46 participants. Of the 111 contacted people, 49 people responded (44%) to the first Delphi round. All 111 invited in the first Delphi round were also invited to participate in the second Delphi round, including 16 others who were recommended as experts but not contacted due to fact that data saturation was reached in the first round. A total of 53 people out of the 127 responded (42%) to the second Delphi round. During the third Delphi round, 49 participants from over 12 different countries responded (92%) and were included in the analysis. Out of the 49 participants in third Delphi round, 30 (61%) had a degree in physiotherapy, 4 (5%) had a degree in exercise physiology or exercise therapy and 14 (29%) had a medical doctor degree. Twenty-nine (59%) participants had a PhD, 41 (84%) worked in academics or a research institute, 5 (10%) in a hospital or an institution, 2 (4%) in private practice or a clinic and 1 (2%) was emeritus.

Stage 1: generating an item pool

The first Delphi round resulted in an item pool of 61 different items based on the comments of 49 experts (see online supplemental appendix 1 for an overview of all 61 items including their scores).

Stage 2: item selection

Out of the 61 available items, 17 were left after applying the cut-off value (table 2). The item 'It is essential for the potential effectiveness of a therapeutic exercise programme to be ethically sound' was the only item with an IQR of 0 and a median of 7. Six other items had a median of 7, while the other 10 items had a median and a 25th percentile score of 6.

Stage 3: item specification

The focus group discussion demonstrated that several items (item 1, 2, 4, 10, 11, 13, 14, 16) (table 2) were multi-interpretatable, which prompted a discussion about how they should be changed. The items were systematically discussed and changed if full consensus of all participants was achieved. Out of the 17 items, all items were suggested to be rephrased and one to be removed (item 9). Two clusters were created, both containing four items to be rephrased into one single item. The transcripts of the focus group are available at request by contacting the first author.

The second author collated the results from the focus group. Based on the suggestions from the focus group, the first and second authors created a survey containing the suggestions and the proposed final items. The participants accepted the changes to the items 1, 2, 4, 11, 14 and 16. The participants accepted both of the clusters, the items 1–4 and 5, 12, 14, 17 and the rephrasing of the items. As a result of the focus group discussion, it was suggested that item 6 would be redundant, as it is already inherent in the new definition of rationale. Therefore, it would have no added value and should be removed. Removal was accepted by all but one of the participants.

Stage 4: developing and refining of the tool

Working group discussion

Before the working group discussion, the first and second author used the results of both the Delphi rounds and the

Table 2 Items left after applying the cut-off value, including median and IQR scores (Q₁ and Q₃)

It is essential for the potential effectiveness of a therapeutic exercise programme:	Median	Q ₁	Q ₃
1. To be based on a plausible rationale.	6	6	7
2. To have a rationale for the mode of exercise.	6	6	7
3. To have a rationale for the dosage of the exercise programme.	7	6	7
4. To have anatomical, physiological, psychological and behavioural relevance to the injury/condition in question.	7	6	7
5. That the content of the exercise programme is related to the goals to achieve.	7	6	7
6. To have the potential to achieve the identified goals.	7	6	7
7. That the mode of exercise is in line with the purpose of the exercise programme.	6	6	7
8. To yield only minimal adverse events.	6	6	7
9. To be ethically sound.	7	7	7
10. That therapy adherence is adequate.	6	6	7
11. That the eligibility criteria select patients that are in need of treatment.	6	6	7
12. To match the goal of the therapeutic exercise to the patients problems.	6	6	7
13. That, in case the exercise programme is supervised, the supervisor's competences and skills are matched to the goals and content of the programme.	6	6	7
14. That the outcome measures reflect the goals of the intervention.	7	6	7
15. That outcomes are assessed with validated performance measures.	7	6	7
16. That outcomes are assessed directly after the intervention.	6	6	7
17. That the outcomes of the exercise programme are explained, based on a plausible rationale.	6	6	7

working group to rephrase the 17 items. The current state of the items were statements, making rephrasing to the criteria the first stage. During the rephrasing, it was noted that the prior established categorisations did not seem applicable or logical. Therefore, a new categorisation has been applied to the items (table 3) selected by the first and second author based on the results and comments from both the focus group and the survey. The produced concept was sent to the members of the working group before the discussion took place via email to collect points of discussion.

The working group discussion contained 5 points based on both the survey, as well as the feedback on the concept. Due to differences in time zones, 4 out of 8 participants were able to attend the working group session. During the discussion, consensus was reached on removing item 6, rephrasing the adherence to the exercise programme, using a high and low risk while not using unclear as an option, and usage of the Frequency, Intensity, Timing, and Type (FITT) criteria for mode and dosage.⁵ The working group concluded that item 9, 'to be ethically sound', had little to no influence on the potential effectiveness of a trial and should therefore not be included in the tool. Changes were applied and sent to all participants for their final commentary, as well as the opinions from the participants who were unable to attend.

Two researchers tested the concept of the tool, independently of each other, on seven different articles. All sections were

deemed necessary without tedious overlap when using the tool. No changes were made.

Checklist items

The final items included in the i-CONTENT tool (see table 4) are: (1) patient selection, (2) dosage of the exercise programme, (3) type of the exercise programme, (4) qualified supervisor, (5) type and timing of outcome assessment, (6) safety of the exercise programme and (7) adherence to exercise programme. The items are briefly described in the text and addressed in detail in the table 4.

Patient selection: When scoring this item, the question at hand is: Were the right patients selected in the study? Meaning that the problems or disabilities of the patient population align with the purpose of the exercise therapy programme. For example, if the goal of an exercise intervention was to improve functional capacity, did the participants selected for this programme have a limited functional capacity?

Dosage of the exercise programme: When scoring this item, the question at hand is: Was it likely that the dosage of the exercise intervention could have resulted in the expected treatment response? A plausible rationale regarding the benefits of the therapeutic exercise programme—especially if there is little or no previous experience with the intervention—is thought to be necessary to achieve therapy effects. The lack of a sound rationale for the dosage of the exercise therapy programme may result in underdosing or overdosing. For example, if the purpose of an exercise programme is to improve functional mobility of frail older adults, did the authors come up with a plausible or proven rationale for dosing the exercise intervention?

Type of exercise intervention: When scoring this item, the question at hand is: Did the type of the exercises match with the purpose of the exercise programme? Type of exercise is defined as the form in which the exercise is provided. In case there is a discrepancy between the type and purpose of the exercise therapy programme, there could be a lack of exercise specificity, which is thought to result in a lower quality exercise programme. For example, if the purpose of an exercise programme is to improve walking capacity, did the authors indeed test a programme

Table 3 The categories included in the tool with their respective original item number

Category	Original item no
1. Patient selection	11,12
2. Dosage	1,3,4,6
3. Type	2,5,7
4. Qualified supervisor	13
5. Type and timing of outcome assessment	14,15, 16,17
6. Safety	8
7. Adherence	10

Table 4 The final version of the i-CONTENT tool**1. Patient selection**

Discrepancy between the problems or disabilities of the patient population and the purpose of exercise therapy programme may result in suboptimal effects.

'Low risk' of ineffectiveness*	<p>The purpose of the exercise therapy programme matches the patients' problems (directly or through a plausible causative relationship). In this case patients' problems can lie in the International Classification of Functioning, Disability and Health domains of body functions, body structures, and activities and participation. For example:</p> <ul style="list-style-type: none"> ▶ If the purpose of the exercise programme is to improve a patient's functional status, then only patients with reduced functional status should be included (direct relationship). ▶ If the purpose of the exercise programme is to improve a patient's balance, then specifically patients with impaired balance need to be included in the study (direct relationship). ▶ If the purpose of the exercise programme is to improve a patient's cardiovascular fitness in patients who have been deconditioned as a result of long-term hospitalisation (indirect relationship).
'High risk' of ineffectiveness	<p>The purpose of the exercise therapy programme does not match the patients' problems. For example: If the purpose of the exercise programme is to improve a patient's quadriceps strength, but patients were not selected on having low quadriceps strength, nor is quadriceps strength a plausible target for the indexed disease</p>

2. Dosage of the exercise programme

The lack of a sound rationale for the dosage of the exercise therapy programme to achieve the purpose of the exercise programme may result in underdosing or overdosing.

'Low risk' of ineffectiveness*	<p>The investigators applied a plausible or proven rationale based on anatomical, physiological, psychological, neurological, or behavioural relevance to the condition to determine the: Frequency\dagger, Intensity\ddagger, and Time\S of the exercise programme matching the purpose of the exercise intervention. For example:</p> <ul style="list-style-type: none"> ▶ Training above 80% of 1RM with low volume for 8 weeks to improve muscle hypertrophy; ▶ Repeating the same movement in high volume to achieve neurophysiological adaptation; ▶ Exercising at least three times a week using resistance training to increase muscular strength.
'High risk' of ineffectiveness	<p>The investigators did not use a plausible or proven rationale based on anatomical, physiological, psychological, neurological or behavioural relevance to the condition or did not match the purpose of the exercise programme. The investigator did use a plausible or proven rationale based on anatomical, physiological, psychological, neurological or behavioural relevance to the condition, but there is a disconnect between the rationale and the applied Frequency, Intensity, and Time of the exercise programme. For example:</p> <ul style="list-style-type: none"> ▶ Primarily using cardiovascular exercise to increase muscle hypertrophy. ▶ Primarily exercising above 80% of 1 RM with low frequency to improve stability. ▶ Primarily exercising 1 day a week at 30% HRR to increase muscular strength, using resistance training.

3. Type of the exercise programme

Discrepancy between the type and purpose of the exercise therapy programme may lead to a lack of exercise specificity.

'Low risk' of ineffectiveness*	<p>The investigators applied a plausible or proven rationale based on anatomical, physiological, psychological, neurological or behavioural relevance to the condition to determine the: Type\P. Furthermore, the investigators matched the Type of the exercise therapy programme with the purpose of the exercise therapy programme. Type of exercise is defined as the form in which the exercise is provided. According to the training specificity principle, it is more likely for training benefits to be transferred to activities if the Type of exercise relates to functional movements. For example:</p> <ul style="list-style-type: none"> ▶ Cardiovascular exercise through running, cycling or dancing. ▶ Resistance exercise through open/closed chain exercises related to functional targets. Furthermore, if an exercise therapy programme is built up from a no of single exercises, also consider whether there is a plausible or proven rationale for each of the exercises.
'High risk' of ineffectiveness	<p>The investigators did not match the type of the exercise programme with the purpose of the exercise therapy programme. For example:</p> <ul style="list-style-type: none"> ▶ Primarily iso-kinetic exercise is prescribed, while the purpose of the programme was walking. ▶ Exercise is primarily delivered by use of exercise bike, while the purpose of the exercise programme was to improve patient-specific activities of daily living.

4. Qualified supervisor (if applicable)

Supervisor(s) who lack the right skills and experiences regarding the exercise programme and patient population may result in suboptimal effects.

Note: In case an exercise intervention was not supervised, forgo scoring this item.

'Low risk' of ineffectiveness*	<p>It can be assumed that the supervisor providing the exercise therapy programme is experienced with the targeted patient population and is sufficiently skilled in providing the proposed exercise programme. For example:</p> <ul style="list-style-type: none"> ▶ If the content of the evaluated exercise therapy programme falls within the basic skill set of a registered physiotherapist, then the necessary skills can be assumed if the programme is delivered by a physiotherapist. However, the experience with the target population would still need to be made clear. ▶ If the content of the evaluated exercise therapy programme falls outside the basic skill set of a registered physiotherapist, the training and experience in the applied programme need to be defined.
'High risk' of ineffectiveness	<p>It can be assumed that the supervisor providing the programme is inexperienced with the patient population or is insufficiently skilled to provide the exercise programme. For example:</p> <ul style="list-style-type: none"> ▶ The exercise programme is provided by a medical specialist without specific training. ▶ The exercise programme is provided to patients aged 25 and below by a medical specialist experienced in geriatric patients.

5. Type and timing of outcome assessment

Using invalid outcome measures or mistiming the measurements might result in the (erroneous) conclusion that an exercise programme was not effective.

Continued

Table 4 Continued

'Low risk' of ineffectiveness*	The investigators used one or more performance-based outcome measures which reflect the goals and purpose of the exercise programme to assess the effectiveness exercise therapy programme. The measurements from the performance-based outcome measures have taken place within the time window where the expected effect would most likely take place. These performance-based measures need to be valid for the targeted patient population as well as for detecting change over time. For example: ▶ The goal of the exercise programme is to improve functional mobility, and the effect of the exercise programme is measured using the Timed Up and Go test; this test can be assumed to be valid for the targeted patient population and is responsive to measuring change over time. Moreover, the timing of the measurements fall within the period in which the optimal effect can be expected.**
'High risk' of ineffectiveness	The investigators use a non-validated (performance) measure as primary outcome measure to assess the effect of the therapeutic intervention. For example, if: ▶ The performance measure applied has not been validated or is unlikely to be valid for the targeted patient group. ▶ The performance measure is known to have a floor or ceiling effect. ▶ The effect of the exercise programme was measured solely by use of self-report measures. ▶ The measurements were obtained outside the window where the optimal effects would be expected.
6. Saafety of the exercise programme A high risk for or no of adverse events may result in a high drop-out rate, reduced adherence and suboptimal effects.	
'Low risk' of ineffectiveness*	The no and severity of the exercise-related adverse events in the study are in line with the expected no of adverse events for similar exercise programmes in similar populations.
'High risk' of ineffectiveness	The no and severity of the exercise related adverse events are substantially higher than what would be expected, possibly resulting in a higher drop-out or reduced level of adherence.
7. Adherence to the exercise programme Low exercise therapy adherence by the patient to the programme may result in a suboptimal effect.	
'Low risk' of ineffectiveness*	Based on relevant information regarding to exercise adherence (ie, the no of sessions attended, the no of exercises performed, and whether or not the intended exercise dosage was reached) the rater draws the conclusion whether the intended exercise dosing was achieved. In order to warrant a 'low-risk' conclusion, the level of adherence of patients to the exercise therapy programme is deemed sufficient to assume that the proposed exercise therapy programme was performed as originally intended, in terms of achieved exercise intensity. Cut-off scores may be used to determine whether adherence was deemed adequate, ⁵⁵ however, we want to stress that the decision needs to be made whether the intended exercise dosing was achieved.
'High risk' of ineffectiveness	The level of exercise adherence of patients to the exercise therapy programme was insufficient to assume the intended exercise programme was performed as intended.

*In case insufficient information is provided to judge this item definitively as 'low risk', then it is up to the rater to make a (conservative) judgement as well as provide a rationale for this judgement.

†The number of days per week dedicated to the exercise programme.

‡Intensity can be defined using several different measures including but not limited to: percentage of maximal oxygen consumption, oxygen consumption reserve, heart rate reserve, maximal heart rate, or metabolic equivalents.⁵

§A measure of amount of time physical activity is performed or by the total caloric expenditure.

¶A variety of exercises to improve the components of physical fitness.

**Both the type of the outcome measurement as well as the timing of the outcome measurement should be reasonable to be able to have a 'low risk'. If either of the two lacks in either reporting or rationale a high risk of ineffectiveness should be assumed.

HRR, heart rate reserve; RM, repetition maximum.

that included walking-type exercises? If the aim of the exercise programme is to improve general well-being, the authors might have selected less specific types of exercises.

Qualified supervisor: When scoring this item, the question at hand is: If a person was supervising the exercise programme, was this person sufficiently qualified? Unsupervised exercise programmes are thought to be of lower quality than supervised programmes. However, the qualities of a supervisor are also thought to influence treatment effects, as supervisors who lack the right skills, experiences and competences regarding both the content of exercise programme as well as the patient population might insufficiently apply an exercise intervention. Depending on the complexity of an exercise intervention and the patient population, the needed qualifications may vary. For example, if a high intensity exercise intervention is assessed in a population of frail older adults with Parkinson's disease, did the authors select supervisors with proven expertise on both the programme and the population?

Type and timing of outcome assessment: When scoring this item, the question at hand is: Is it likely that the treatment response to the exercise intervention was actually measured? To adequately measure the response to an exercise intervention, it is important that a measurement tool is valid and responsive,

but also that the measurement tool was deployed at the right moment in time. All three elements are thought to be of importance to avoid drawing erroneous conclusions. For example, if the purpose of an exercise programme is to increase physical activity by stimulating participants to slowly increase their own exercise regimens at home, did the authors measure the physical activity with a valid tool and at the right timing?

Safety of the exercise programme: When scoring this item, the question at hand is: Is the exercise programme safe? It is thought that an exercise programme with a high-risk adverse events related to the intervention may result in a high drop-out rate and/or reduced adherence, which might result in inflated and/or suboptimal effects. The risk for skewed outcomes because of adverse events should be contemplated. For example, if a high intensity exercise programme is administered to frail older adults, did patients drop-out with adverse events (resulting in selective reporting) and/or did patients (and supervisors) deviate from the intended treatment protocol?

Adherence to the exercise programme: When scoring this item, the question at hand is: Did the patients adhere to the exercise programme as it was described in the methods section? Insight into adherence is relevant as low exercise therapy adherence by the patient to the programme is thought to result in

suboptimal effects. For example, if an exercise intervention aims to make people with severe obesity be more physically active by requiring them to perform 150 min of moderate activity per day, did the patients adequately adhere to this programme?

Checklist scoring

People using the tool are required to judge each item as either: 'low risk' or 'high risk' for ineffectiveness as well as provide a rationale to support their judgement. If the information was not explicitly reported in the manuscript, the reviewers are required to provide a judgement and a rationale to support their judgement (to this end, we included 'probably done' and 'probably not done' to the scoring sheet (see online supplemental appendix 2). The wording on the two judgement criteria and the scoring sheet match those of the Cochrane's Risk of Bias tool.³³ In line with the Cochrane's Risk of Bias tool, no overall score should be calculated, but each item should be weighted on their importance within the study that is assessed (ie, quality of the single study) and in unison with all other studies that are assessed (ie, the body of evidence). We suggest a narrative assessment be made on the therapeutic quality at an individual study level and on the total body of evidence (ie, all studies combined).

We recommend people systematically reviewing the literature on exercise therapy should assess both risk of bias of the included studies as well as quality of the studied interventions and interpret these outcomes in conjunction. After all, poor methodological quality of the used study design can inflate study outcomes,^{34–38} which might erroneously be interpreted as a superior exercise intervention. Finally, we recommend that a reviewer who rates the quality of the exercise intervention be blinded towards the outcomes of this study.

CONCLUDING REMARKS

As the number of scientific publications on—and the number of prescriptions for—exercise therapy continue to grow, we believe a better understanding of the quality and content of these interventions in the scientific literature is warranted. We believe the i-CONTENT tool will be a starting point for researchers, healthcare professionals, and peer reviewers to take intervention quality into account and move the exercise therapy evidence base to the next level. While further validation is necessary, this can be done by the exercise therapy community by critically applying the i-CONTENT and refining the instrument in parallel.

The i-CONTENT tool represents a considerable expansion over previous efforts to elucidate the quality of exercise interventions. The current approach has the primary aim to create a rating tool, rather than a reporting guideline.^{9 18–21} We believe the size and composition of the Delphi group, containing a range of experts from 12 different countries, lend credibility to the tool. Moreover, our rigorous approach to collate and group the Delphi items into the final seven items helped create an unbiased tool. A final strength is that the seven items of the tool are all supported by scientific evidence. Several studies have shown that proper patient selection influences the effectiveness of treatment due to differences in responses, potentially leading to greater therapy gains.^{39–42} The impact of both dosage and type of the exercise programme on its effectiveness due to the direct dose response relationship has been well established in the literature.^{43 44} Also, qualified supervisor (in terms of acquired skills and experience) is known to influence the treatment effects, for example, due to the increased adherence when treated by a trained professional.^{45–47} In that same line of reasoning, safety of the therapy can be important, as a high risk of adverse events

may result in high drop-out rates, reduced adherence and suboptimal effects.^{48 49} Furthermore, the validity of the instrument to measure the response to the exercise intervention, as well as the timing and its frequencies of that measurement, can impact an intervention's effectiveness.¹⁸ Finally, to ensure if the prescribed dosage has been performed, adherence to the exercise programme has to be maintained and appropriately described.^{50–53}

There are a number of limitations to our work. First, we did not include a patient-representative in the working group. Second, a Delphi panel with a different composition might have resulted in a somewhat different tool.⁵⁴ As we primarily focused on exercise therapy, other professions including exercise physiology and sports medicine, might not have been well represented by our panel. On the other hand, exercise scientists were part of the group, and data saturation was reached for the initial Delphi round, suggesting that contacting more experts would not have led to different input. Furthermore, the decision to reject or accept items, was made on an arbitrary level of importance. Nevertheless, we feel that both the working group, as well as the Delphi panel were sufficiently knowledgeable concerning the essential ingredients that make up high-quality exercise interventions. Moreover, the level used to select items was set a priori and was consistent with previous studies. Finally, to provide full transparency in which items were in- and which were excluded, a detailed list with all specific scores is provided in online supplemental appendix 1.

We developed a tool to assess the therapeutic quality of RCTs. We hope that i-CONTENT tool for short, will result in better health and (physical) functioning of patients via prevention and care concepts stemming from improved exercise therapies. The tool may also help researchers and clinicians gain new insights in exercise therapy due to a better understanding of the current body of evidence and may set a new standard for the quality of RCTs. The i-CONTENT tool will be dynamic in its nature, as new insights will help shape the content and composition/structure of the tool over time.

What are the findings?

- ▶ The international Consensus on Therapeutic Exercise and Training (i-CONTENT) tool is a step towards transparent assessment of the quality of exercise therapy programmes studied in randomised clinical trial. The tool adds to the existing reporting guidelines, as it structures the weighing, interpretation, and value of the relative potential of exercise therapy to possess the theoretical and practical potential to improve a person's (physical) functioning.

How might it impact on clinical practice in the future?

- ▶ The i-CONTENT tool provides clinicians and researchers a resource to better identify, appraise and interpret the heterogeneity across trials of exercise, and ultimately, to assist in the development of future, higher quality, exercise interventions.

Author affiliations

¹Radboud Institute for Health Sciences, IQ healthcare, Radboudumc, Nijmegen, Gelderland, The Netherlands

²Biomedical Sciences, Radboud University, Nijmegen, Gelderland, The Netherlands

³Executive Director, Top Sector Life Sciences & Health (Health-Holland), The Hague; Professor, Dept Anesthesiology, Erasmus MC, Rotterdam; CEO, Topcare, The Netherlands

⁴Global Aging, Cochrane Collaboration, London, UK

⁵Department of Sports Medicine, Norwegian School of Sports Sciences, Oslo, Akershus, Norway

⁶Department of Obstetrics and Gynecology, Akershus University Hospital, Lorenskog, Norway

⁷Department of Medicine, University of Ottawa, Ottawa, Ontario, Canada

⁸Institute of Bone and Joint Research, The Kolling Institute, University of Sydney Sydney Medical School, Sydney, New South Wales, Australia

⁹Department of Epidemiology, Maastricht University, Maastricht, Limburg, The Netherlands

¹⁰Department of Rheumatology, Sint Maartenskliniek, Nijmegen, Gelderland, The Netherlands

¹¹Department of Physical Medicine and Rehabilitation, University of Colorado Denver School of Medicine, Aurora, Colorado, USA

¹²Eastern Colorado VA Geriatric Research Education and Clinical Center (GRECC), Aurora, Colorado, USA

Correction notice This article has been corrected since it published Online First. The affiliations for Prof van Meeteren have been updated.

Acknowledgements Special thanks to Bart Staal and Ria Nijhuis-van der Sanden for attending the focus group to give their independent opinion on the proposed subjects and to Jordi Elings (MSc) for helping consolidate the 17 items into the final 7seven items. Appreciations also go to Jane Green from Caledonian University for testing the concept and exchanging experiences on using the concept of the tool.

Contributors All authors comply with recommendations for contributorship by the ICMJE: substantial contributions to the conception or design of the work; or the acquisition, analysis or interpretation of data for the work; and drafting the work or revising it critically for important intellectual content; and final approval of the version to be published; and agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as online supplemental information.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Thomas J Hoogeboom <http://orcid.org/0000-0003-2103-419X>

Kari Bo <http://orcid.org/0000-0003-1176-9272>

Peter Tugwell <http://orcid.org/0000-0001-5062-0556>

REFERENCES

- Huber M, Knottnerus JA, Green L, *et al*. How should we define health? *BMJ* 2011;343:d4163.
- Naci H, Ioannidis JPA. Comparative effectiveness of exercise and drug interventions on mortality outcomes: metaepidemiological study. *BMJ* 2013;347:f557.
- Gates AB, Kerry R, Moffatt F, *et al*. Movement for movement: exercise as everybody's business? *Br J Sports Med* 2017;51:767–8.
- Pedersen BK, Saltin B. Exercise as medicine - evidence for prescribing exercise as therapy in 26 different chronic diseases. *Scand J Med Sci Sports* 2015;25:1–72.
- ACSM. *ACSM's Guidelines for Exercise Testing and Prescription*. Lippincott Williams And Wilkins, 2013: 480.
- Beauchamp TL, Childress JF. *Principles of biomedical ethics*, 2013.
- Herbert RD, BQ K. Analysis of quality of interventions in systematic reviews. *BMJ* 2005;331:507–9.
- Dagfinrud H, Halvorsen S, Vøllestad NK, *et al*. Exercise programs in trials for patients with ankylosing spondylitis: do they really have the potential for effectiveness? *Arthritis Care Res* 2011;63:597–603.
- Hoogeboom TJ, Oosting E, Vriezekolk JE, *et al*. Therapeutic validity and effectiveness of preoperative exercise on functional recovery after joint replacement: a systematic review and meta-analysis. *PLoS One* 2012;7:e38031.
- Jo D, Del Bel MJ, McEwen D, *et al*. A study of the description of exercise programs evaluated in randomized controlled trials involving people with fibromyalgia using different reporting tools, and validity of the tools related to pain relief. *Clin Rehabil* 2019;33:557–63.
- Neil-Sztramko SE, Medysky ME, Campbell KL, *et al*. Attention to the principles of exercise training in exercise studies on prostate cancer survivors: a systematic review. *BMC Cancer* 2019;19:321.
- Abell B, Glasziou P, Hoffmann T. Reporting and replicating trials of exercise-based cardiac rehabilitation: do we know what the researchers actually did? *Circ Cardiovasc Qual Outcomes* 2015;8:187–94.
- Tew GA, Brabyn S, Cook L, *et al*. The completeness of intervention descriptions in randomised trials of supervised exercise training in peripheral arterial disease. *PLoS One* 2016;11:e0150869.
- Holden S, Rathleff MS, Jensen MB, *et al*. How can we implement exercise therapy for patellofemoral pain if we don't know what was prescribed? A systematic review. *Br J Sports Med* 2018;52:385.
- Knols RH, Fischer N, Kohlbrenner D, *et al*. Replicability of physical exercise interventions in lung transplant recipients; a systematic review. *Front Physiol* 2018;9:946.
- Slade SC, Finnegan S, Dionne CE, *et al*. The consensus on exercise reporting template (CERT) applied to exercise interventions in musculoskeletal trials demonstrated good rater agreement and incomplete reporting. *J Clin Epidemiol* 2018;103:120–30.
- Neil-Sztramko SE, Winters-Stone KM, Bland KA, *et al*. Updated systematic review of exercise studies in breast cancer survivors: attention to the principles of exercise training. *Br J Sports Med* 2019;53:504–12.
- Schulz KF, Altman DG, Moher D, *et al*. Consort 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;340:c332.
- Chan A-W, Tetzlaff JM, Altman DG, *et al*. Spirit 2013: new guidance for content of clinical trial protocols. *The Lancet* 2013;381:91–2.
- 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.
- Slade SC, Dionne CE, Underwood M, *et al*. Consensus on exercise reporting template (CERT): modified Delphi study. *Phys Ther* 2016;96:1514–24.
- Therapeutic validity of exercise therapy in RCTs. Puijo symposium; 2014; Kuopio.
- Mokkink LB, Terwee CB, Patrick DL, *et al*. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 2010;19:539–49.
- Verhagen AP, de Vet HC, de Bie RA, *et al*. The Delphi list: a criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. *J Clin Epidemiol* 1998;51:1235–41.
- Murphy E, Dingwall R, Greatbatch D. Qualitative research methods in health technology assessment: a review of the literature. *Health Technol Assess* 1998;2:iii–ix. 1–274.
- Jensen GM, Gwyer J, Shepard KF, *et al*. Expert practice in physical therapy. *Phys Ther* 2000;80:28–43. discussion 44–52.
- Altschuld JW, Thomas PM. Considerations in the application of a modified scree test for Delphi survey data. *Eval Rev* 1991;15:179–88.
- Powell C. The Delphi technique: myths and realities. *J Adv Nurs* 2003;41:376–82.
- Francis JJ, Johnston M, Robertson C, *et al*. What is an adequate sample size? Operationalising data saturation for theory-based interview studies. *Psychol Health* 2010;25:1229–45.
- Yates SL, Morley S, Eccleston C, *et al*. A scale for rating the quality of psychological trials for pain. *Pain* 2005;117:314–25.
- Corporation R. R-05 wave/mp3 recorder owner's manual, 2010. Available: https://www.roland.com/us/support/by_product/r-05/owners_manuals/9e574bcf-0259-40c0-b872-fed4a1904181/
- Alexandrov AV, Pullicino PM, Meslin EM, *et al*. Agreement on disease-specific criteria for do-not-resuscitate orders in acute stroke. *Stroke* 1996;27:232–7.
- Higgins JPT, Altman DG, Gotzsche PC, *et al*. The Cochrane collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;343:d5928.
- Schulz KF, Chalmers I, Altman DG. The landscape and lexicon of blinding in randomized trials. *Ann Intern Med* 2002;136:254–9.
- Schulz KF, Grimes DA. Blinding in randomised trials: hiding who got what. *The Lancet* 2002;359:696–700.
- Schulz KF, Chalmers I, Hayes RJ, *et al*. Empirical evidence of bias. dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;273:408–12.
- Schulz KF, Grimes DA, Altman DG, *et al*. Blinding and exclusions after allocation in randomised controlled trials: survey of published parallel group trials in obstetrics and gynaecology. *BMJ* 1996;312:742–4.

- 38 Gluud LL. Bias in clinical intervention research. *Am J Epidemiol* 2006;163:493–501.
- 39 Hoeksma H *et al.* Manual therapy in osteoarthritis of the hip: outcome in subgroups of patients. *Rheumatology* 2005;44:461–4.
- 40 McGinn TG, Guyatt GH, Wyer PC, *et al.* Users' guides to the medical literature: XXII: how to use articles about clinical decision rules. *JAMA* 2000;284:79–84.
- 41 Veenhof C, Van den Ende CHM, Dekker J, *et al.* Which patients with osteoarthritis of hip and/or knee benefit most from behavioral graded activity? *Int J Behav Med* 2007;14:86–91.
- 42 Wright AA, Cook CE, Flynn TW, *et al.* Predictors of response to physical therapy intervention in patients with primary hip osteoarthritis. *Phys Ther* 2011;91:510–24.
- 43 Courneya KS, McKenzie DC, Mackey JR, *et al.* Effects of exercise dose and type during breast cancer chemotherapy: multicenter randomized trial. *J Natl Cancer Inst* 2013;105:1821–32.
- 44 Gallois M, Davergne T, Ledinot P, *et al.* Dosage of preventive or therapeutic exercise interventions: review of published randomized controlled trials and survey of authors. *Arch Phys Med Rehabil* 2017;98:2558–65. e10.
- 45 Resnik L, Jensen GM. Using clinical outcomes to explore the theory of expert practice in physical therapy. *Phys Ther* 2003;83:1090–106.
- 46 Boutron I, Moher D, Altman DG, *et al.* Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. *Ann Intern Med* 2008;148:295–309.
- 47 Hawley-Hague H, Boulton E, Hall A, *et al.* Older adults' perceptions of technologies aimed at falls prevention, detection or monitoring: A systematic review. *Int J Med Inform* 2014;83:416–26.
- 48 Basaria S, Coviello AD, Travison TG, *et al.* Adverse events associated with testosterone administration. *N Engl J Med* 2010;363:109–22.
- 49 Bouchard C, Blair SN, Church TS, *et al.* Adverse metabolic response to regular exercise: is it a rare or common occurrence? *PLoS One* 2012;7:e37887.
- 50 Hayden JA, van Tulder MW, Tomlinson G. Systematic review: strategies for using exercise therapy to improve outcomes in chronic low back pain. *Ann Intern Med* 2005;142:776–85.
- 51 Campbell *Ret al.* Why don't patients do their exercises? Understanding non-compliance with physiotherapy in patients with osteoarthritis of the knee. *J Epidemiol Community Health* 2001;55:132–8.
- 52 Miller FL, O'Connor DP, Herring MP, *et al.* Exercise dose, exercise adherence, and associated health outcomes in the tiger study. *Medicine & Science in Sports & Exercise* 2014;46:69–75.
- 53 Kruger C, McNeely ML, Bailey RJ, *et al.* Home exercise training improves exercise capacity in cirrhosis patients: role of exercise adherence. *Sci Rep* 2018;8:99.
- 54 Campbell SM, Hann M, Roland MO, *et al.* The effect of panel membership and feedback on ratings in a two-round Delphi survey: results of a randomized controlled trial. *Med Care* 1999;37:964–8.
- 55 Osho O, Owoeye O, Armijo-Olivo S. Adherence and attrition in fall prevention exercise programs for community-dwelling older adults: a systematic review and meta-analysis. *J Aging Phys Act* 2018;26:304–26.