European Journal of Physical and Rehabilitation Medicine 2024 Sep 18
DOI: 10.23736/S1973-9087.24.08361-8

SPECIAL ARTICLE

Randomized controlled trials as a source of evidence in rehabilitation: a critical analysis

Sandra SCHMITZ *, Thorsten MEYER-FEIL

Interdisciplinary Center of Health Sciences, Institute for Rehabilitation Medicine, Faculty of Medicine, Martin Luther University Halle-Wittenberg, Halle, Germany

*Corresponding author: Sandra Schmitz, Interdisciplinary Center of Health Sciences, Institute for Rehabilitation Medicine, Faculty of Medicine, Martin Luther University Halle-Wittenberg, Magdeburger Straße 8, 06112 Halle, Germany. E-mail: sandra.schmitz@uk-halle.de

ABSTRACT

The randomized controlled trial (RCT) is the study design with the greatest potential to maximize internal validity when assessing the effectiveness of medical interventions, making it invaluable for evidence-based medicine. Yet, especially in the field of rehabilitation, it is not universally accepted as an unassailable gold standard due to serious problems of its implementation. This paper first examines three factors that limit the applicability of RCTs in rehabilitation practice. The first two factors stem from the nature of rehabilitative treatment itself: the complexity of rehabilitation interventions and the long-term and holistic nature of rehabilitation goals. The third factor relates to the differing functions of RCTs. Interventions vary in their complexity in increasing degree between component, measure, and program interventions. Lower complexity is associated with a greater likelihood of using high rigor efficacy studies. Methodological rigor further depends on the degree to which intervention conditions or contexts can be controlled for. This is particularly the case when examining body-related short-term outcomes. Whether it is reasonable to conduct an RCT also hinges on its function: to gain knowledge or to legitimate the utilization of an intervention in rehabilitation practice. The discussion highlights key challenges to RCT implementation and states questions that should help to identify an RCT as the most appropriate research design. Further empirical and theoretical research is indicated to clarify the distinction between levels of intervention, as this paper is based on theoretical considerations. Additionally, a concise explication of the different functions of an RCT and its meanings for their implementation is needed.

(*Cite this article as:* Schmitz S, Meyer-Feil T. Randomized controlled trials as a source of evidence in rehabilitation: a critical analysis. Eur J Phys Rehabil Med 2024 Sep 18. DOI: 10.23736/S1973-9087.24.08361-8)

KEY WORDS: Rehabilitation; Rehabilitation research; Evidence-based medicine; Randomized controlled trial; Methods.

Introduction

The evidence-based medicine (EBM) or, more broadly, evidence-based practice (EBP) movement represents an impressive success story in health care. EBM looks at health-related research from the perspective of clinical or health care decision-making: which scientific evidence can or should inform practitioners' decisions? The idea itself seems to match general expectations of both clinicians and patients perfectly, as Dickersin *et al.* pinpointed: "It is curious, even shocking, that the adjective 'evidence based' is needed. The public must wonder on what basis medical decisions are made otherwise.… Is it intuition? Magic?"[1] EBM was strongly influenced by ideas developed to test the efficacy of medical drugs, and its methodology is rooted in approaches from clinical epidemiology. However, now it is indispensable in all areas of health care. Its approaches have been applied even beyond health care, *e.g.* to politics (see Strassheim *et al.*[2] for a critical view).

The EBM movement also resonated early (*e.g.*, Lae *et*

al.[3]) and thoroughly in the field of rehabilitation. At the heart of these endeavors should be the rehabilitation patient with his or her individual problems, needs, and rehabilitation goals. A high degree of individualization of treatment plans and the associated multimodal treatment approaches, as well as collaboration in multi- or interprofessional teams, are the basic prerequisites for achieving these goals. The use of interventions, for which effectiveness has been demonstrated, is of concern not only for rehabilitation professionals who set up intervention plans and provide interventions, but also for other stakeholders:[4] rehabilitation patients, family caregivers, funding bodies or policymakers.[5]

However, questions and unease persist regarding the way evidence is produced, transferred, and used in rehabilitation.[6-8] This was one central motivation for the founding of Cochrane Rehabilitation in 2016, which aims to bridge the well-known evidence gap between the spheres of rehabilitation research and practice.[9] One object of criticism is the perceived one-to-one adoption of the gold standard for intervention studies, the randomized-controlled trial (RCT).

The RCT is still considered the best study design for conducting a summative evaluation of an intervention in order to be able to attribute measurable effects exclusively to that intervention. By means of randomization and blinding, adequately implemented RCTs allow researchers to control for and minimize risks of biases and confounding. Thus, the internal validity of study results is maximized. It is the most appropriate design to rule out alternative explanations of differences between intervention and control group. In rehabilitation research, the potential of a high internal validity of this study type remains undisputed. The feasibility of RCTs, however, especially blinding and randomization, has been disputed repeatedly.[6, 10] A higher internal validity of the study, which might be reached by high levels of standardization of treatment procedures, is related to artificial study conditions. This reduces the potential of transferring the study results into practice, and therefore the usefulness of the study.[10, 11] Also, the limited scope and severe limitations in the selection of the study population, which lead to issues of limitations in external validity and clinical replicability have been a source of criticism.

Within an evidence-based medicine framework, the design of a study is both affected by methodological standards (researchers' side) and the needs of the users of evidence, especially clinicians (users' side). The potential tension between different requirements is expressed by the differentiation of levels of evidence and grade of recommendations within the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach.[12] It seems clear that in designing a study, researchers should select the most optimal methodological approach in accordance with the research question and research interest in order to ensure that the study is able to answer the research question. From the perspective of the users of evidence a pragmatically designed and conducted RCT might be preferable to a more methodologically rigorous approach, whereby the quality of the evidence is not determined by the theoretically optimal study design, but by pragmatic needs for evidence determined by potential users.

It seems that the potential of the study design in terms of a scientific-theoretical ideal does not necessarily translate into a reasonable implementation in research practice[8] and that the results of these studies lack clinical relevance.[6] In their scoping review on methodological issues in rehabilitation research, Arienti *et al.*[13] listed both problems related to the conduct of a study and its respective reporting that relate to threads of internal and external validity. It became apparent that the majority of issues that were identified in the review were not intrinsic to the design of RCTs but rather related to the way they were conducted or reported.

There are issues that refer to problems that are beyond the scope of conduct or reporting and relate instead to challenges that arise from the intrinsic characteristics of rehabilitation. For example, behavioral interventions dominate in rehabilitation practice, and these can hardly be reduced to singular cause-effect relationships and transferred to larger populations without multiple assumptions both about the way the intervention is delivered and about variable context factors.[6, 14]

Opposition to the use of RCTs in rehabilitation might also have something to do with the peculiarities that make up rehabilitation. Especially within the last 15 years different attempts have been made to frame a common understanding of rehabilitation worldwide.[15] It should be worthwhile to reflect on these approaches as possible sources to explain the unease parts of the rehabilitation practice community has in the idea of using evidence from RCTs.

To sum up, the scientific discussion surrounding the role of RCTs as a source of evidence for rehabilitation practice is still not settled. This paper aims to advance this central discussion by providing important theoretical distinctions based on current conceptual developments on the definition of rehabilitation and a theoretical reflection on different functions that RCTs serve in rehabilitation practice.

These distinctions, presented as a methodological note, should be regarded as central to the appraisal of the relevance of different types of RCTs for an evidence-based rehabilitation practice.

Methodological considerations

The present paper is based on theoretical reflections which relate to the preliminary work of Cochrane Rehabilitation, especially on a scoping review on methodological issues in rehabilitation research.[13] Therefore, we developed our line of arguments based on the conceptual and theoretical conclusions of other researchers and our prior theoretical work which was also inspired by the long-standing conceptual work of one of the authors (TMF) on the definitions of rehabilitation[16] and the involvement in controversial discussions within Cochrane Rehabilitation related to RCTs.[17, 18] The authors of the aforementioned scoping review have designated this type of work as "theoretical thinking,"[13] *i.e.*, articles used to explain a theoretical concept or issue without a systematic inclusion of primary studies.

Characteristics of rehabilitation

Various attempts to define rehabilitation have already been made, representing different perspectives.[15] Accordingly, rehabilitation could be seen as a health strategy, a process, or a set of measures while its "specific aim is the optimization of aspects of functioning, especially social participation, independence or self-determination (which are actually related but can be very different aims), or quality of life."[15] A recent definition of rehabilitation has been developed by Cochrane Rehabilitation in a thorough and comprehensive multistage process involving different professional stakeholder groups in the field.[18] This definition uses the PICO-framework and characterizes rehabilitation within the context of health care as follows: "In a health care context, rehabilitation is defined as a multimodal, person-centered, collaborative process including interventions targeting a person's capacity (by addressing body structures, functions, and activities/participation) and/or contextual factors related to performance with the goal of optimizing the functioning of persons with health conditions currently experiencing disability or likely to experience disability, or persons with disability."[18]

Therefore, rehabilitation is a comprehensive approach for people in different health conditions, while each rehabilitation patient should receive a specific combination of measures tailored to his or her needs and aligned with personal goals. As a result of individual tailoring, a rehabilitation patient is always treated with multimodal therapy approaches, ideally by a multidisciplinary team. These approaches include, among others, physical exercises, behavioral or psychological measures, education, counselling, and the provision of devices and assistive technologies.[19] These measures potentially affect multiple rehabilitation goals, which may change and be adapted in the progress of rehabilitation, are based on the patient's individual perspectives and are the result of a shared decision-making process between a clinician and the rehabilitation patient. These goals might be distinguished into short-term *vs.* long-term goals, or, put in a similar way, rehabilitation (or process)-related *vs.* personal goals. While short-term, process-related goals are often referred to aspects of physical or mental capacities (body functions),[20] long-term goals relate to performance in everyday life, such as return-to-work or being able to supervise children, or other goals related to aspect of participation or to an individual`s quality of life.[21]

In contrast to short-term goals, which should be achievable during rehabilitation and therefore need to be discussed with the treating therapists ("negotiated goals"[20]), long-term goals are closely linked to the patient's basic values, inner attitude and worldview ("meaningful overall goals"[21]). Short-term goals are therefore only intermediate steps on the way to achieving long-term goals, which are of particular importance to the person undergoing rehabilitation. These goals should be in close correspondence with the choice in outcomes in RCTs.

In almost every case, behavioral therapy approaches play a central role. In pharmacological treatments, there is an external agent (medication) that does the work. In contrast, rehabilitation measures work by an inextricable interaction between the therapist and the patient, usually but not necessarily with support of some technology.[6] Their success depends on the professional competencies of the therapists (cf. REREP[22]) but also on "the attention and action of the person[s]" involved and include their "motives, values, and thoughts,"[23] just as with any other interventions with educational and psychosocial features. This implies that the potential effect of these interventions is substantially dependent on factors that may initially remain elusive to the treating person and might become apparent only during the rehabilitation process, if at all. This also implies that healthcare professionals in rehabilitation resort to many unspecific competencies, motivations, attitudes, and skills to build up a positive relationship with

the patient. Therefore, most rehabilitation measures are complex as they involve these interactive processes. The complexity in these measures might be described best "as a dynamic and constantly emerging set of processes and objects that not only interact with each other, but come to be defined by those interactions."[24] As these interactions are very individual, even standardized measures are expected to vary widely in practice.

Given the definition of rehabilitation described above, individual measures can only be considered as part of a rehabilitation process, not as rehabilitation itself. The reason for this can be derived from the rehab-cycle model that integrates assessment, goal-setting, assignment, intervention, and evaluation based on the functional health model of the International Classification of Functioning, Disability and Health (ICF).[25, 26] The output of rehabilitation care (the service) should be seen as the result of a collaborative effort between provider(s) and the patient.[18]

This perspective has important consequence regarding the characteristics of the intervention in a rehabilitation study. We should distinguish different levels of rehabilitation that could be considered as interventions in a study (Supplementary Digital Material 1: Supplementary Text File 1).[18-20, 27-30]

A rehabilitation program consists of available measures, the combination of which is suitable (based on internal and/or external evidence) to achieve meaningful overall rehabilitation goals. Such a program can be, for instance, cardiac rehabilitation after a heart attack. Rehabilitation patients and therapists can decide on an individualized combination from suitable and available measures. The offer of rehabilitation programs depends in particular on the available services of the health care system and the respective structures (accessibility, financing, guidelines, common minimum standards). The term "rehabilitation program" is thus used as a synonym to the term "rehabilitation" in the definition of Cochrane Rehabilitation (see above).

Measures are part of the rehabilitation process to achieve individual goals set by patient and therapist. These goals should lead to the achievement of an overall goal of the rehabilitation process and therefore must be oriented towards individual and achievable rehabilitation goals, for example, exercises to increase physical activity. Measures are largely determined by the specific rehabilitation setting, *i.e.*, among other things, by the staffing, the present equipment/technologies, the rehabilitation facility, the costs, etc. Measures can differ in the structure and combination of various components, as well as in the duration, intensity, or frequency applied.

Individual components of measures, such as a specific exercise on a workout machine, are the units of which a measure is composed. A component is characterized by different elements such as duration, frequency, and intensity. An element might be defined by the fact that a variation of it does not affect the other elements. These elements are called "ingredients" in the Rehabilitation Treatment Specification System.[27] We prefer the term "element" over "ingredient" as the former refers to essential parts of a component while the latter could also include optional parts (Supplementary Text File 1).[18-20, 27-30]

These levels of intervention vary in complexity, as do the questions that can be answered using RCTs (see below). Whilst individual components can stand alone, at the level of measures interaction effects must be assumed to occur among multiple components. As mentioned before, personal and environmental factors play an important role in the rehabilitation process ("person-centered"), including motivational and situational factors such as rehabilitation patient's individual environment, which could serve as an important rehabilitation target. Furthermore, the rehabilitation process is also influenced by additional location-specific context factors such as the local healthcare and social systems, which should be taken into account.[31] In fact, the sensitivity of clinical studies to context factors might be understood as the step taken from the clinical study to health-services research studies, the primary interest of which is context effects.[32] Effective rehabilitation is therefore dependent on various characteristics and contextual factors determined by the setting and the individual needs of each person undergoing rehabilitation.

In conclusion, rehabilitation is much more than a single measure. It is complex, and its complexity increases depending on which part of the process is being looked at in a study context: a component of a measure, a measure within the rehabilitation process, or the set of measures within a rehabilitation process, *i.e.*, a rehabilitation program.

Nevertheless, despite this complexity, it remains essential to gain a deeper understanding of the effectiveness and underlying mechanisms of rehabilitation and its constituent parts, especially if the aim of research is to be able to appraise the benefits of an intervention.

The important distinction between efficacy and effectiveness trials

When we look at RCTs, it is reasonable to distinguish between two types of studies: efficacy trials gather evidence

on what works under ideal conditions in a more strictly selected population, whereas effectiveness trials are intended to support evidence of the effect of an intervention under usual conditions of clinical care in a more broadly defined clinically relevant population.[6, 33] The two types of study are therefore based on different aims. Efficacy trials aim at answering the question if an intervention works at all. They use an artificial, almost mechanistic approach to try to establish a causal link between a specific intervention and a subsequent effect by keeping contextual factors constant.[34] It, therefore, focuses on a high level of internal validity at the expense of elements of external validity. Effectiveness trials are designed to prove that an intervention is effective despite allowing for contextual factors to vary (in fact, these are modelled as error variance, not as explanatory variables). While it therefore focusses on aspects of external validity, it is in our view not appropriate to say that this should be at the expense of internal validity, as internal validity can be considered as a prerequisite of external validity. Interventions should be proven to be effective in an efficacy trial with high level of internal validity prior to analyzing their effectiveness that aim to allow for broad generalizations to usual clinical care. Both study types have in common that they are designed to demonstrate a causal effect.[34, 35]

General problems with the application of RCTs

If the mechanisms behind the effect of an intervention on the outcome being studied are known, it must also be known that these mechanisms are valid outside the study conditions being investigated.[36] In particular, if mechanisms are unknown, various factors contributing to the outcome besides the intervention studied may not be addressed.[36] Therefore, it is important to consider whether the study aims at gaining knowledge on causal mechanisms (see below). While these issues are of concern to the study design in general, there are other issues which arise when the study design is to be implemented in practice.

Adequate randomization of participants to study arms might be challenging. RCTs attempt to control bias by random allocation of patients to different groups and therefore to result in a random distribution of (both known and unknown) prognostic factors. Therefore, concealed allocation of study participants to different study groups based on a replicable, but unpredictable, randomization process is essential to successful RCTs.[37] The homogeneous distribution of known and unknown prognostic factors between the study groups is central, since only one condition of one person can be observed at one time, and thus a similar comparison must be used to infer a causal relationship between cause and its assumed effect. However, since an RCT does not look at individual cases but at mean effects of the groups examined, structural equality is established at the group level. Therefore, a predictable randomization sequence, or a non-covert randomization process, leads to an increased risk of systematic bias when study participants are specifically assigned to one group or another based on their personal characteristics. It is important to use a method that guarantees participant concealment, distinguishing between concealment mechanisms and allocation sequence generation.[37] The concealment mechanism utilizing sequentially numbered opaque sealed envelopes is considered particularly unsecure. When this method is chosen, it is crucial to ensure that the recruitment process is separate from the allocation process, and that the two processes are performed by different people. Using small block sizes for a stratified block randomization by site is mostly regarded as an unsafe method for generating sequences. If possible, larger or variable block sizes should be used avoiding stratifying by site.[37] In trials with over 100 participants, it is recommended to use simple randomization to ensure the integrity of the results.

With regard to the target criterion defined as the primary outcome, the required sample size is determined in order to be able to prove a statistically significant effect of an intervention. This is already essential in the planning phase of an RCT in order to avoid *post-hoc* "fishing for significance" in search of any statistically significant effect.[38] Drop-outs, which occur with several follow-up measurements and especially with long follow-up periods,[38] can on the one hand lead to loss of required power that is not achieved in order to keep the probability of a β-error as low as possible. On the other hand, systematically missing values threaten the internal validity of a study. The risk of bias is particularly increased by loss-to-follow-up if the true value of the outcome is a cause of the proportionally higher number of drop-outs in one of the subgroups examined. It is not only for this reason that it is essential to accurately define the outcome of a study.[39, 40]

Essential problems of applying RCTs in rehabilitation research

"[T]he only study design capable of showing a causal relationship between the intervention (rehabilitation) and the outcomes [*i.e.*, the RCT] is not feasible in methodological, ethical or even legal terms".[41] This fundamental statement

refers to the use of RCTs to provide evidence on absolute effectiveness on rehabilitation, *i.e.*, whether a rehabilitation program works compared to no rehabilitation.[42] However, there are rebuttals to this argument. From a methodological point of view, there seems to be no reason why an RCT for the study of absolute effectiveness should not be feasible. The only real methodological problem relates to blinding.[13] Double blinding, *i.e.*, both for patients and clinicians, is mostly impossible in rehabilitation research practice. Only the blinding of raters of outcomes and those doing the analysis of the study data is possible (and worthwhile[43]). Blinding has been considered a central aspect of an RCT because it should control for expectation biases (*e.g.*, Hawthorne and Rosenthal effects[7, 38]).

However, it has also been argued that the blinding of patient and therapist would really distort the true intervention effect, as the expectation ("placebo") effect can be understood as an intrinsic part of the "true" intervention effect (cf. "efficacy paradox"[44]). Also, a so-called "true" intervention effect cannot be identified independent of an expectation effect. Therefore, effectiveness trials should better reflect the true possible effects of rehabilitation measures or programs as they reflect the situation of rehabilitation patients and professionals in rehabilitation practice. The main advantage of an RCT, it can be argued, is the randomization,[45] and this is not distorted by a lack of blinding. Therefore, blinding might be central for the RCT, but is not essential for getting relevant evidence from RCT.[46] Detection bias could be minimized by the use of blinded raters in both efficacy and effectiveness trials.

Another methodological problem relates to the essence of rehabilitation, *i.e.*, the aim of improving long-term functioning with a focus on activities and participation. Therefore, there is always a substantial time difference between the end of the rehabilitation phase and the determination of study outcome, *e.g.*, return-to-work. When the study phase, in which the intervention context can be controlled, ends, for example within an inpatient rehabilitation facility, factors of the home and social environment of the rehabilitation patients can dilute the possible effect of the intervention.

Another essential problem relates to legal terms. If patients have the right to take part in a rehabilitation program, it is rarely possible to randomize those patients that have applied or that have been granted a rehabilitation to a non-rehabilitation group. However, there are study designs that have overcome this problem elegantly, either by screening subjects of possible need for rehabilitation and randomizing them into a rehabilitation counselling group *vs.* no further information group,[42] or by applying a waiting list design.[47] The latter is only possible in cases where the demand or need for rehabilitation exceeds rehabilitation beds considerably, which makes it impossible to provide a rehabilitation for every patient in need to begin with. Also, the follow-up time should not exceed the respective waiting time, which could limit the use of certain long-term outcomes.

Therefore, ethical reasons to refrain from the implementation of RCTs are not convincing. Ethical concerns associated with random allocation of participants to study groups should be addressed by reference to the principle of equipoise.[48] Before planning and implementing an RCT, there must be a balanced uncertainty as to whether one of the applications to be compared within the study arms has a benefit over the other application,[49] which is still often the case. Also, it should be kept in mind that it is ethically questionable to end up with inconclusive study results, *e.g.*, due to a too small sample size.[50]

Important information about clinically relevant long-term rehabilitation outcomes should be based on patient-reported outcomes (PROMS) because it is the patients themselves who are most knowledgeable about what they do and do not do in their daily lives. Short-term rehabilitation goals can also be captured by functional test or observer rated outcomes, such as the six-minute walking test. It could be argued that these PROMS should be automatically prone to detection bias, but this could depend on the framing of the trial and research question for the patients.

Two additional points are also important as possible sources of bias: studies with comparatively large-scaled samples in the sense of several hundreds or even thousands of participants are not often possible, since in rehabilitation research often small populations are involved, for example due to rare diseases or smaller study units within a rehabilitation facility.[4, 23] This may result in smaller effects not being detected or only inaccurately estimated, which in turn can lead to results of these studies only partially reported or not being reported at all (reporting bias). In addition, efficacy trials might not reflect the actual treatment situation, which is a specific threat to external validity.[4] Rehabilitants with comorbidities are excluded regularly in these trials (selection bias) which can also be found to a lesser degree in effectiveness studies.

Another point to consider is a possible blurring of the intervention effect, which can already occur by the fact that the control group of an RCT also receives a treatment, for example in the sense of a treatment-as-usual (TAU) in effectiveness trials, that might even contain components of

the intervention under investigation.[35] It is therefore all the more crucial that the treatment received in both, the intervention and the control group, is described in as much detail as possible.[13, 51] These exact descriptions are also lacking in publications of RCTs, which is a major shortcoming as effectiveness-studies are increasingly conducted.

Functions of RCTs in rehabilitation: stakeholders' interests

For a thorough appreciation of RCTs in rehabilitation, it is worth reflecting on the purposes or function of RCTs as sources of evidence: to gain knowledge on causal mechanisms or to legitimize decisions. Both purposes depend in part on the addressee of the information on evidence. As addressees we consider the patients (and potentially their relatives or family caregivers), the clinicians (the professionals interacting with the patients), and political decision makers (*i.e.*, those persons responsible for managing health care on an institutional level and allocating resources and also those persons in charge of setting guidelines or regulations). Results of RCTs could be generally used to make statements on the probability that a person will benefit from an intervention (with no assurance that this will work in a certain individual case). Here, patients should primarily be interested in the proof that an intervention works, not why the intervention works. This interest comprises in particular whether the interventions the patients take part in do benefit them as an individual, *i.e.*, allow them to better reach their personal (health-related) goals, and at the same time to do no substantial harm (or at least have a favorable benefit-harm balance). For an estimation of harm, RCTs are not well-suited to identify especially low-frequency harms. They are less likely to be detected in RCTs due to limited sample sizes.

It is reasonable to assume that political decision makers have a similar, but broader, perspective on the role of RCTs. They should be interested in evidence that an intervention works and does well than harm but in the overall group of patients, or (sub-)populations. The likelihood of benefit and harm is therefore more important to them than for the individual. This holds true for harms and leads to the need for different study designs to detect low-frequency harms of interventions. Additionally, policymakers tend to have greater interest in studies of cost-effectiveness, which may not be in the primary interest of patients. In conclusion, the purpose of the RCT for this particular group of stakeholders is the legitimization of interventions. It is not primarily about gaining knowledge on causal mechanisms.

For clinicians, the situation can be complex. Firstly, they must act in the best interest of their patients and prioritize evidence of potential benefits from RCTs, including evidence of positive benefit-harm ratios. Secondly, they must also consider society's interests in achieving an adequate cost-effectiveness ratio, as this will be a focus for political decision-makers. Also, it is necessary to present objective evidence to legitimize treatment decisions, which may include consideration of expected effect size as indicated by RCTs. Furthermore, RCTs can serve as a source of knowledge regarding potential causal mechanisms related to the intervention, allowing for adaptation or alteration of treatment decisions.

Essential problems of external validity and generalizability

As with any other study design, the results of an RCT can primarily be related to the population studied and its place and time, because that is where we observed a potential effect. In order to be able to extrapolate results to a larger population, a study population is, in the best case, randomly selected from the target population. This can hardly be the case in rehabilitation, as institutional settings usually each attract a special group of patients, depending on their reputation and services offered. In this sense, studying rehabilitation patients from one clinical setting could be considered as taking a convenience sample.[52] There is at least one study that shows that this problem can be avoided:[42] potential rehabilitation patients are actively contacted and offered rehabilitation. This also gives the investigators a good overview of the non-participants. It can be assumed that studies that include a wider range of settings and a larger study population increase the likelihood that the results of a studied intervention will be of practical use. Especially after major policy changes, it seems useful to re-examine highly complex rehabilitation programs to see whether the existing study results can still be applied to current rehabilitation practice.

Discussion

Designing and conducting high-quality randomized controlled trials can be challenging, even more so for medical rehabilitation, which pursues a holistic, bio-psychosocial approach that goes beyond the bio-medical curative approach. This is already conditioned by the fact that the target group of rehabilitative measures is very diverse and includes patients after acute events, people with chronic

diseases and people with (long-term) disabilities. Rehabilitation involves physical and psychological exercise, education, counselling, and the use of specialized technology. The multimodal interventions that are used may vary in the structure and combination of different components, as well as in the duration, intensity, or frequency with which they are applied. The structure and combination are individualized and tailored. They account for the constitution, motivation, and situational factors of the rehabilitation patient. This results in individual rehabilitation goals, which are also negotiated with the respective therapists, while other actors, such as relatives or funding bodies, also have an interest in the success of rehabilitation activities. Acute events usually require a single, rapid intervention, while chronic diseases require regular therapeutic measures; therapists and rehabilitation patients pursue one or more rehabilitation goals related to the individual, while funding bodies (have to) consider goals or rehabilitation outcomes on a group level; and behavioral or psychological measures require a higher degree of individualization than physical measures. In other words, in considering whether an RCT is suitable for testing the effects of an intervention and whether the methodology is appropriate for carrying it out, the question is not only whether and how the intervention as such can be tested, but also the kind of outcome(s) and the level and context of the intervention. In contrast to many pharmacological studies, where the use of a placebo in the control group means that the cause of a directly observable/measurable effect can be clearly attributed to the drug given due to controllable context conditions, these three points are almost impossible to achieve in rehabilitation. This leads to four overarching problems with RCTs in rehabilitation:

Problem 1: Purpose of the RCT *vs.* methodological adequacy, or does the end always justify the means?

In principle, the aim of the RCT is to establish the causal relationship between the intervention and the change in the measured result. However, we need to look at the purpose of an RCT: should the results of the RCT be used to explain this change (How does it work?) or to legitimize using the intervention in practice (Does it work?)? This is no longer related to the question of whether RCTs can be conducted in rehabilitation in general. Rather, it is a question of where the limits of the use of RCTs lie in relation to these two purposes. In addition, the more complex the intervention is, the less possible it is to pinpoint possible causal mechanisms. Legitimation might then become the most prominent purpose of these studies.

Problem 2: Internal *vs.* external validity

This is a fundamental problem of study design that is even more acute in rehabilitation: in order to make causal inferences, potential confounding factors must be controllable. At the level of the study sample, this is done by randomization. Ideally, confounding factors will be evenly spread by random allocation of participants between study arms. However, the greater the heterogeneity of the individuals, the greater the number of possible confounding factors, so that it becomes increasingly difficult to ensure comparability between the groups. In rehabilitation, it is common to encounter individuals with multiple comorbidities, a variety of social backgrounds, and other contextual factors that are difficult to control, especially over time. Methodologically appropriate conduct of an RCT in the sense of an efficacy trial severely limits the selection of the study population and the observation period, which makes it difficult to extrapolate the study results to the target population.

Problem 3: Intervention interacting with the context/individual situation of the rehabilitation patient

Contextual factors are a problem not only regarding the internal validity of the study results. They are also a problem because of possible interactions with the intervention being studied. If the mechanisms by which the intervention works are unknown, it is also unknown what influence the context of the study might have on the intervention, and thus what role the context might play in the results of the study. Effectiveness trials should therefore be performed by applying the intervention within different contexts, *e.g.*, different clinical settings, that should be representative of the different settings in a specified health care region. Also, it is reasonable to assume that the number of contextual factors, and thus the number of possible interactions, will increase as the observation period gets longer.

Problem 4: Intervention components may interact with each other

Due to the multimodal structure of rehabilitation, no single measure constitutes the entirety of rehabilitation treatment. Therefore, several applications or measures and their components are applied, which can potentially influence each other. Therefore, as described in problem 3, unless the mechanisms of action of the intervention are known, it is unclear how strongly the components or measures influence each other, and thus what proportion of the study results are due to the interaction. The more components and measures the rehabilitation includes, the more likely it

is that the combination of components and measures will have an impact on the study results.

Given the problems outlined above, three essential considerations must be made when using the RCT as a study design:

- What is the purpose of the study?
- At what level (component, intervention, program) should the intervention be classified?
- How rigorous is the realization of an RCT design?

The most methodologically rigorous approach is required for efficacy studies. They can contribute to a gain of knowledge if they are conducted with appropriate rigor. However, because of the many requirements that need to be met, as outlined in the problems, they are mainly suitable for evaluating interventions at the component level. It should also be borne in mind that mainly short-term outcomes can be measured, in order to exclude the uncontrollable influence of contextual factors, especially in the home environment.

Interventions at the program level are somewhat different. As they involve a wide variety of contextual factors, it makes sense to conduct them as an effectiveness trial with a pragmatic approach. It is important to bear in mind that an interest in a gain of knowledge in terms of identifying causal mechanisms cannot be served. Due to the factors mentioned above, a legitimation interest dominates in these cases. It is also possible to evaluate long-term outcomes, but one must be aware that possible effects cannot be attributed to the rehabilitation intervention from a scientific perspective.

A differentiated view is necessary if interventions are to be evaluated at the measure level. Provided that the mechanisms of action and the influence of the interaction of the included components on the outcome are known, it is also possible to conduct efficacy trials at this level. However, this applies to proximal outcomes. If long-term outcomes are to be assessed, or if the mechanisms of action are not known, effectiveness trials are the first choice.

Conclusions

In this paper we have worked out the special features in the field of rehabilitation that make it difficult to conduct randomized controlled trials in a methodologically adequate way. However, we have also shown that this is possible and useful if certain conditions are met. Noteworthy are the importance of the levels of intervention and the different purposes of RCTs. In this context, and this can be seen as a limitation, our elaborations are based on the literature known to us as well as on thought experiments. In further empirical and theoretical work, the differentiation and individual components of the intervention levels should be worked out more clearly. Also, the purposes of an RCT should be more clearly stated, *e.g.*, by means of a review, and the differences between the purposes mentioned should be highlighted.

References

1. Dickersin K, Straus SE, Bero LA. Evidence based medicine: increasing, not dictating, choice. BMJ 2007;334(Suppl 1):s10.

2. Strassheim H, Kettunen P. When does evidence-based policy turn into policy-based evidence? Configurations, contexts and mechanisms. Evid Policy 2014;10:259–77.

3. Law MC, editor. Evidence-based rehabilitation: a guide to practice. Thorofare, NJ: Slack; 2002.

4. Carter RE. Lubinsky J. Rehabilitation research: principles and applications. Fifth Edition. St. Louis, MO: Elsevier; 2016.

5. Stucki G, Reinhardt JD, Grimby G. Organizing human functioning and rehabilitation research into distinct scientific fields. Part II: conceptual descriptions and domains for research. J Rehabil Med 2007;39:299–307.

6. Hart T, Bagiella E. Design and implementation of clinical trials in rehabilitation research. Arch Phys Med Rehabil 2012;93(Suppl):S117–26.

7. Kersten P, Ellis-Hill C, McPherson KM, Harrington R. Beyond the RCT - understanding the relationship between interventions, individuals and outcome - the example of neurological rehabilitation. Disabil Rehabil 2010;32:1028–34.

8. Negrini S. Evidence in rehabilitation medicine: between facts and prejudices. Am J Phys Med Rehabil 2019;98:88–96.

9. Negrini S, Kiekens C, Levack W, Grubisic F, Gimigliano F, Ilieva E, *et al.* Cochrane physical and rehabilitation medicine: a new field to bridge between best evidence and the specific needs of our field of competence. Eur J Phys Rehabil Med 2016;52:417–8.

10. Fuhrer MJ. Overview of clinical trials in medical rehabilitation: impetuses, challenges, and needed future directions. Am J Phys Med Rehabil 2003;82(Suppl):S8–15.

11. Tate DG, Findley T Jr, Dijkers M, Nobunaga AI, Karunas RB. Randomized clinical trials in medical rehabilitation research. Am J Phys Med Rehabil 1999;78:486–99.

12. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, *et al.*; GRADE Working Group. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008;336:924–6.

13. Arienti C, Armijo-Olivo S, Minozzi S, Tjosvold L, Lazzarini SG, Patrini M, *et al.* Methodological issues in rehabilitation research: a scoping review. Arch Phys Med Rehabil 2021;102:1614–1622.e14.

14. Johnston MV, Dijkers MP. Toward improved evidence standards and methods for rehabilitation: recommendations and challenges. Arch Phys Med Rehabil 2012;93(Suppl):S185–99.

15. Meyer T, Kiekens C, Selb M, Posthumus E, Negrini S. Toward a new definition of rehabilitation for research purposes: a comparative analysis of current definitions. Eur J Phys Rehabil Med 2020;56:672–81.

16. Meyer T, Gutenbrunner C, Bickenbach J, Cieza A, Melvin J, Stucki G. Towards a conceptual description of rehabilitation as a health strategy. J Rehabil Med 2011;43:765–9.

17. Levack WM, Meyer T, Negrini S, Malmivaara A; Cochrane Rehabilitation Methodology Committee. Cochrane Rehabilitation Methodology Committee: an international survey of priorities for future work. Eur J Phys Rehabil Med 2017;53:814–7.

18. Negrini S, Selb M, Kiekens C, Todhunter-Brown A, Arienti C, Stucki G, *et al.*; 3rd Cochrane Rehabilitation Methodology Meeting participants. Rehabilitation definition for research purposes. A global stakeholders' initiative by Cochrane Rehabilitation. Eur J Phys Rehabil Med 2022;58:333–41.

19. World Health Organization. World Bank. World Report on Disability 2011. Geneva: WHO; 2011.

20. Preede L, Soberg HL, Dalen H, Nyquist A, Jahnsen R, Saebu M, *et al.* Rehabilitation goals and effects of goal achievement on outcome following an adapted physical activity-based rehabilitation intervention. Patient Prefer Adherence 2021;15:1545–55.

21. Dekker J, de Groot V, Ter Steeg AM, Vloothuis J, Holla J, Collette E, *et al.* Setting meaningful goals in rehabilitation: rationale and practical tool. Clin Rehabil 2020;34:3–12.

22. Negrini S, Arienti C, Pollet J, Engkasan JP, Francisco GE, Frontera WR, *et al.*; REREP study participants. Clinical replicability of rehabilitation interventions in randomized controlled trials reported in main journals is inadequate. J Clin Epidemiol 2019;114:108–17.

23. Johnston MV, Sherer M, Whyte J. Applying evidence standards to rehabilitation research. Am J Phys Med Rehabil 2006;85:292–309.

24. Cohn S, Clinch M, Bunn C, Stronge P. Entangled complexity: why complex interventions are just not complicated enough. J Health Serv Res Policy 2013;18:40–3.

25. WHO. International Classification of Functioning, Disability and Health: ICF. WHO: Geneva; 2001.

26. Negrini S. Physical and rehabilitation medicine. J Int Soc Phys Rehabil Med 2019;2:S47–54.

27. Van Stan JH, Whyte J, Duffy JR, Barkmeier-Kraemer JM, Doyle PB, Gherson S, *et al.* Rehabilitation Treatment Specification System: methodology to identify and describe unique targets and ingredients. Arch Phys Med Rehabil 2021;102:521–31.

28. Krug E, Cieza A. Strengthening health systems to provide rehabilitation services. Bull World Health Organ 2017;95:167.

29. Rocca E, Anjum RL. Causal Evidence and Dispositions in Medicine and Public Health. Int J Environ Res Public Health 2020;17:1813.

30. ICHI Task Force. ICHI beta-3 reference guide: International Classification of Health Interventions; 2020 [Internet]. Available from: https://icd.who.int/dev11/l-ichi/en [cited 2024, Sep 3].

31. Gutenbrunner C, Nugraha B. Decision-making in evidence-based practice in rehabilitation medicine: proposing a fourth factor. Am J Phys Med Rehabil 2020;99:436–40.

32. Meyer T. Evidenzbasierung der Rehabilitation [Evidence based rehabilitation]. In: Meyer T, Bengel J, Wirtz MA, editors. Lehrbuch Rehabilitationswissenschaften [Textbook rehabilitation sciences]. Hogrefe: Bern; 2022. p.294–306. [German].

33. Whitty CJ. What makes an academic paper useful for health policy? BMC Med 2015;13:301.

34. Flay BR, Biglan A, Boruch RF, Castro FG, Gottfredson D, Kellam S, *et al.* Standards of evidence: criteria for efficacy, effectiveness and dissemination. Prev Sci 2005;6:151–75.

35. Gottfredson DC, Cook TD, Gardner FE, Gorman-Smith D, Howe GW, Sandler IN, *et al.* Standards of evidence for efficacy, effectiveness, and scale-up research in prevention science: next generation. Prev Sci 2015;16:893–926.

36. Cowen N, Virk B, Mascarenhas-Keyes S, Cartwright N. Randomized controlled trials: how can we know 'what works'? Crit Rev 2017;29:265–92.

37. Clark L, Fairhurst C, Torgerson DJ. Allocation concealment in randomised controlled trials: are we getting better? BMJ 2016;355:i5663.

38. Shadish WR, Cook TD, Campbell DT. Experimental and quasi-experimental designs for generalized causal inference. Belmont, CA: Wadsworth Cengage Learning; 2002.

39. Vetter TR, Mascha EJ. Defining the primary outcomes and justifying secondary outcomes of a study: usually, the fewer, the better. Anesth Analg 2017;125:678–81.

40. Küçükdeveci AA, Tennant A, Grimby G, Franchignoni F. Strategies for assessment and outcome measurement in physical and rehabilitation medicine: an educational review. J Rehabil Med 2011;43:661–72.

41. Gerdes N, Zwingmann C, Jäckel W. The system of rehabilitation in Germany. In: Jäckel W, Barth J, Herdt J, editors. Research in rehabilitation: results from a research network in southwest Germany. Stuttgart: Schattauer;2006:3–19.

42. Hüppe A, Langbrandtner J, Lill C, Raspe H. The effectiveness of actively induced medical rehabilitation in chronic inflammatory bowel disease. Dtsch Arztebl Int 2020;117:89–96.

43. Moustgaard H, Clayton GL, Jones HE, Boutron I, Jørgensen L, Laursen DR, *et al.* Impact of blinding on estimated treatment effects in randomised clinical trials: meta-epidemiological study. BMJ 2020;368:l6802.

44. Zhang W, Doherty M. Efficacy paradox and proportional contextual effect (PCE). Clin Immunol 2018;186:82–6.

45. Stephenson J, Imrie J. Why do we need randomised controlled trials to assess behavioural interventions? BMJ 1998;316:611–3.

46. Malmivaara A, Armijo-Olivo S, Dennett L, Heinemann AW, Negrini S, Arokoski J. Blinded or nonblinded randomized controlled trials in rehabilitation research: a conceptual analysis based on a systematic review. Am J Phys Med Rehabil 2020;99:183–90.

47. Schultz K, Wittmann M, Wagner R, Lehbert N, Schwarzkopf L, Szentes B, *et al.* In-patient pulmonary rehabilitation to improve asthma control: a randomized controlled study (EPRA, Effectiveness of Pulmonary Rehabilitation for Patients with Asthma). Dtsch Arztebl Int 2021;118:23–30.

48. Beauchamp TL, Childress JF. Principles of biomedical ethics. New York, NY: Oxford University Press; 2019.

49. Miller FG, Joffe S. Equipoise and the dilemma of randomized clinical trials. N Engl J Med 2011;364:476–80.

50. Halpern SD, Karlawish JH, Berlin JA. The continuing unethical conduct of underpowered clinical trials. JAMA 2002;288:358–62.

51. Whyte J, Dijkers MP, Fasoli SE, Ferraro M, Katz LW, Norton S, *et al.* Recommendations for Reporting on Rehabilitation Interventions. Am J Phys Med Rehabil 2021;100:5–16.

52. Dijkers MP. External validity in research on rehabilitative interventions: issues for knowledge translation; 2024 [Internet]. Available from: https://ktdrr.org/ktlibrary/articles_pubs/ncddrwork/focus/ [cited 2024, Sep 3].

Conflicts of interest
The authors certify that there is no conflict of interest with any financial organization regarding the material discussed in the manuscript.

Funding
The authors acknowledge the financial support for publication costs from the Open Access Publication Fund of Martin Luther University Halle-Wittenberg.

Authors' contributions
Sandra Schmitz and Thorsten Meyer-Feil contributed equally to the manuscript. Both authors read and approved the final version of the manuscript.

History
Article first published online: September 18, 2024. - Manuscript accepted: September 2, 2024. - Manuscript revised: July 15, 2024. - Manuscript received: December 1, 2023.

Box 1: Specification of terms

We use the term *measures* to refer to action packages related to one modality, profession, or specific rehabilitation target applied within the rehabilitation process. The use of this term refers to the definition of rehabilitation by WHO and the World Bank in the World Report on Disability [19]: "a set of measures that assist individuals who experience, or are likely to experience, disability to achieve and maintain optimal functioning in interaction with their environments" (p. 308). We could use the term *intervention* interchangeably (as the WHO has done in a subsequent definition in 2017 [28]), but restrict it in this paper to denote only specific manipulations within the study context, which are actions that only the intervention group receives.

Components of measures are the smallest units that could be isolated and used for itself or in a different context and bear the possibility for change in the patients' functioning. We use the term *component* as understood within the RTSS-framework in its current version by Stan et al. [27].A professional should not be regarded as a component. However, professional experience or specific training of professionals are important aspects, which, once defined, should be considered as preconditions for the (proper) implementation of interventions within a study. From a dispositionalist's perspective [29], these preconditions (or dispositions) are nevertheless essential for achieving an intended effect of the intervention being studied.

A rehabilitation *programme* is composed of different measures. The programme is characterised by a set of different options in terms of measures, that can (but do not have to) be applied based on the individual patient's needs and goals. It relates to the definition of rehabilitation as explicated by Cochrane Rehabilitation, i.e., to the understanding of rehabilitation as a process that relates to the goal of-optimisation of functioning.

The term *goal* is kept general here. The focus is on the goals of the rehabilitation patients, which "should reflect the perspective of the persons living with the disability" (p. 1545) [20]. In an intervention study, however, the goals of other stakeholders, such as insurers, may also be relevant, and intersections will inevitably arise. It is crucial that goals define a desired future (functional) state as a result of a process. In rehabilitation practice, the rehabilitation patient and a professional should set goals in mutual agreement. A distinction is made between short- and long-term goals.

A *target* differs from a goal, as it refers to the entity on which an action is carried out. In reference to the definition of rehabilitation by Cochrane Rehabilitation this could include "a person's capacity (by addressing body structures, functions, and activities/participation) and/or contextual factors related to performance" (p. 336) [18]. We use this term in accordance with its definition within the model of the International Classification of Interventions (p. 6) [30]. The use of the term *target* there is different from its use within the RTSS framework, where it is related to level of functioning to be achieved [27].