# Faithful but flexible: Intervention Fidelity in Clinical Trials of Complex interventions in Healthcare

Arsenio Páez

A DPhil Thesis
Submitted to the Department of Continuing Education
University of Oxford,
In Fulfilment of the Requirements for
The Doctor of Philosophy (DPhil) in Evidence-based Healthcare
The University of Oxford
March 2023

## Abstract

The focus of clinical trials is typically interventions' efficacy, or whether they attain their desired outcomes. Comparatively less attention is focused on understanding how or why interventions succeed, or fail to attain, those outcomes. This may be particularly important in trials of complex interventions such as surgery or physiotherapy, which are multifaceted and often tailored to individual participants, providers, or settings, increasing the potential for variations in intervention delivery and effects.

The correspondence between the intervention that was planned and what was actually delivered in a trial is the intervention's fidelity. Several benefits for high levels of intervention fidelity have been proposed. However, whether or how much fidelity influences clinical trials' treatment effect estimates had not previously been determined. A lack of a uniform definition for fidelity and its key components may also hinder intervention delivery in clinical trials and the translation of evidence-based interventions to clinical practice.

The principle aim of this thesis was to investigate the effects of intervention fidelity on the results, interpretation, and appropriateness to change practice of clinical trials of complex interventions. Through systematic review, "Best-fit" framework synthesis, reliability study, meta-epidemiological study, and fidelity assessment in an on-going pragmatic randomised controlled trial (RCT), it addressed several important knowledge gaps for intervention fidelity in complex intervention clinical trials. It estimated the prevalence of fidelity monitoring and reporting in

complex intervention RCTs. It contributed an empirically-based intervention fidelity framework and a reliable, internally consistent fidelity assessment checklist tailored to the unique needs of researchers who seek to assess clinical trials in rehabilitation. It also estimated empirically, for the first time, the magnitude and direction of bias in treatment effect estimates arising from poor intervention fidelity in rehabilitation RCTs. Fidelity assessment in the ACL-SNNAP trial also provided new insights into the feasibility, applicability, and impact of monitoring intervention fidelity and participant adherence in complex intervention pragmatic trials.

The findings of this thesis showed that intervention fidelity is important, not just on a theoretical, but also on an empirical level. This thesis provides new and important information for the conduct and interpretation of clinical trials in rehabilitation and other complex interventions in healthcare.

## Acknowledgements

Alfred North Whitehead said: “No one who achieves success does so without acknowledging the help of others.” There are countless people who could be thanked for their contributions, large or small, toward the completion of this DPhil thesis. Among these, I am particularly grateful to my primary DPhil supervisor and mentor, Prof. David Beard. I have been, and continue to be, fortunate for his support, encouragement, generosity, and expertise. I am deeply grateful that he constantly challenged and inspired me to see further and dig deeper into the nature and life of things. This DPhil emerged from a conversation with Prof. Beard and I will always be grateful for the great and transformational privilege it has been for me to work with him. It was the first step in an exciting and continuing journey. Thank you, Prof. Beard for that first nudge onto the road!

I would be hard-pressed to express the depths of my gratitude to Dr. David Nunan, who was an incredible source of ideas, energy, enthusiasm, and insight during my DPhil. I stand in awe of his expertise and knowledge and am deeply grateful for his guidance during this journey. I am also deeply grateful to Prof. Dan Hind, at the University of Sheffield, for his great kindness, generosity, enthusiasm, and expertise. It has been my privilege to learn from, and work with, Prof. Hind, without whom I could not have accomplished some of the most important components of this thesis. I am also deeply grateful to Prof. Peter McCulloch, whose wisdom, guidance, knowledge, and support were invaluable for my DPhil and my development as a researcher and academic.

I extend my sincere thanks for the guidance and support I received from the tutors and staff at the Evidence-Based Healthcare Program, the Nuffield Department for Primary-Care Health Sciences, Centre for Evidence-based Medicine, the Department for Continuing Education, and Kellogg College. I found a welcoming and vibrant community and a home at Kellogg College. I am a proud and grateful Kelloggian. I will always be grateful to the Evidence-based Healthcare program for the transformational experiences I have had in both the MSc and DPhil in EBHC. Thank you, Dr. Plüdermann and Prof. Heneghan.

Finally, and with great love and affection, I wish to thank and acknowledge my husband Steven for his support and encouragement throughout my DPhil. It was a life-long dream for me to read for a DPhil at the University of Oxford, and one finally realized later in life. Thank you for your support and encouragement, and your many, nameless, unremembered (by you, but not by me) acts of kindness and of love. To further paraphrase Wordsworth: After many wanderings, many years, this experience has been dear to me, both for itself, and for your sake! Thank you.

## Chapters overview

### Chapter I: Introduction:

An introduction and overview of the thesis. The history and evolution of intervention fidelity and current evidence-gaps are discussed. Overviews of the chapters and methods in the thesis are presented.

### Chapter II: CONSIDER: A “best-fit” framework synthesis

A systematic review quantifies and describes intervention fidelity frameworks, models, and concepts reported in papers of complex interventions in the physical domain (surgery, physiotherapy, and rehabilitation) and follows a best-fit framework synthesis methods to develop the Complex Interventions Design, Delivery, Receipt (CONSIDER) a conceptual framework and checklist, and integrated definition for intervention fidelity for this thesis.

### Chapter III: Testing the reliability and consistency of the CONSIDER checklist

The CONSIDER checklist and its explanation and elaboration paper and scoring guide are described. The checklists’ internal consistency, reliability, and inter-rater reliability are investigated in a reliability study.

### Chapter IV: Intervention fidelity and treatment effect estimates in complex intervention trials.

A meta-epidemiological study of randomized clinical trials in complex interventions is undertaken to determine whether intervention fidelity biases the magnitude and/or direction of treatment effect estimates derived from RCTs of physical complex interventions.

## Chapter V: The ACL-SNNAP trial

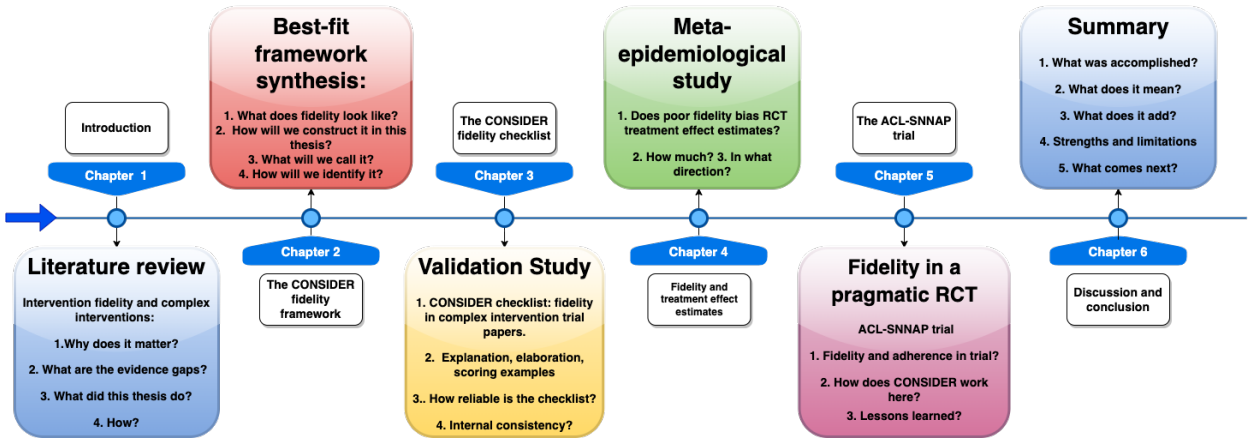
This chapter gathers insights about the CONSIDER framework and the results of the meta-epidemiological study through the greater level of granularity available from real-time data gathered prospectively in an ongoing clinical trial, the Anterior Cruciate Ligament (ACL) Surgery Necessity in Non-Acute Patients (ACL-SNNAP) trial. Key aspects of the framework and the meta-epidemiological study in Chapter IV were validated and several important insights about fidelity monitoring and fidelity reporting in complex intervention clinical trials were gained.

## Chapter VI: Discussion and conclusions

The results and implications of the thesis, along with its strengths, limitations, and future directions are discussed.

## Chapter VII: Appendices

This section includes search strategies and results, checklists, and supplemental tables.

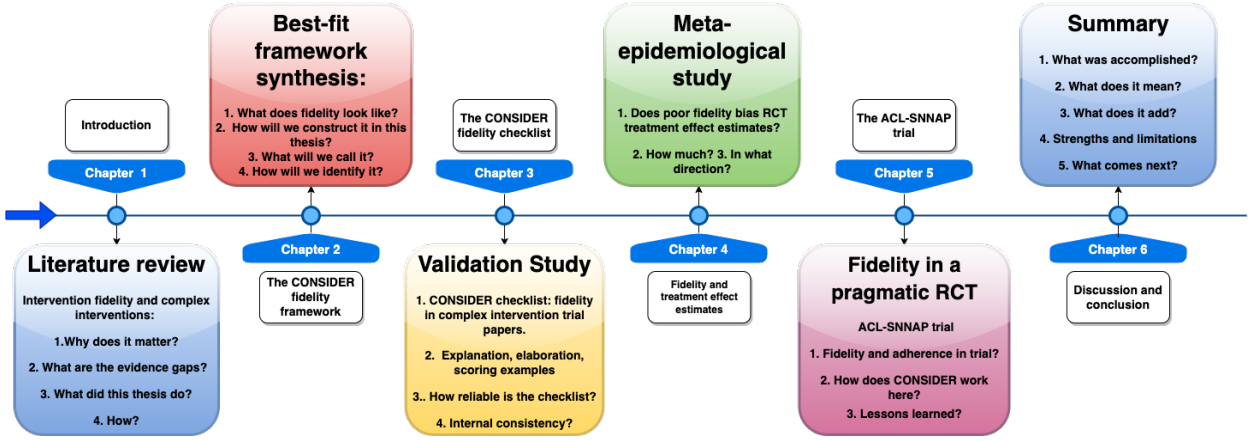



Thesis flow diagram

# Chapter I: Introduction and background: Intervention fidelity and complex interventions

## Chapter Summary

**Chapter I** offers an introduction to this thesis. It also presents the rationale and importance of investigating how intervention fidelity affects the results of clinical studies of complex interventions, their interpretation, and their appropriateness for guiding changes in clinical practice. It presents the history and development of intervention fidelity monitoring in clinical trials and efforts to improve evidence and innovation in complex interventions in healthcare. A systematic literature review of intervention fidelity in complex interventions reveals major gaps in the evidence-base for intervention fidelity in complex interventions, laying the context for the thesis. The chapter then describes the structure and main aims of the thesis and gives a description of the general methods used.



## INTRODUCTION

Outside of safety, the focus of clinical trials is typically intervention efficacy, or whether or not interventions attain their desired outcomes. By comparison, less attention is focused on understanding how or why interventions succeed, or fail to attain, their target outcomes.[1,2] This may be of particular importance in trials of complex interventions.[3–7] Complex interventions are multifaced, highly individualised, and context dependent.[3–5,8,9] Complex interventions can be difficult to standardize and their delivery can vary greatly between participants and settings, posing unique challenges for clinical investigations of intervention effects.[3–7]

Intervention fidelity generally refers to the faithfulness of the intervention delivered in a clinical trial to the intervention that was intended in the trial protocol. Several benefits from high levels of intervention fidelity and negative consequences arising from poor intervention fidelity have been proposed. However, a number of overlapping terms and constructs are used to describe and operationalise intervention fidelity, hampering efforts to support and monitor it in clinical trials.[10–12] Additionally, whether poor intervention fidelity has a quantifiable influence on intervention effect estimates derived from complex intervention clinical trials has not been investigated, nor has the direction of bias arising from poor intervention fidelity been determined. Whether poor intervention fidelity enhances or neutralises intervention effects has not yet been determined or shown empirically.

This thesis investigated the effects of intervention fidelity on the results of clinical studies of complex interventions, their interpretation, and their appropriateness for guiding changes in

clinical practice. It offers a conceptual fidelity framework and fidelity checklist for clinical trials of complex interventions in physical domains in healthcare, such as physiotherapy and rehabilitation. It quantifies, for the first time, the effect of intervention fidelity on treatment effect estimates (the estimated causal effect of the experimental intervention [independent variable] on the primary outcome [dependent variable] of the study) derived from RCTs of complex interventions, with important implications for the interpretation of systematic reviews and meta-analyses of complex intervention RCTs and clinical practice guidelines based on them. This thesis also applies the conceptual fidelity framework to investigate intervention fidelity and adherence in detail in a contemporary pragmatic randomised trial, exploring the effect of fidelity violations on the interpretation of the trial's results and the applicability of the conceptual framework to pragmatic trials.

## Complex interventions

Complex interventions, such as such as surgery and physiotherapy, involve a number of components which may act independently and interdependently to achieve some desired end.[5,7,13,14] Practitioner skill and experience, learning curves, provider-patient expectations, differences in individual patient characteristics, biopsychosocial factors, and other factors can contribute to the outcome of complex interventions,[5–7] sometimes antagonistically.[15] For example, the effectiveness of surgical interventions may be enhanced, or degraded, by surgeons' experience with the procedure, the degree to which procedure components have been defined, patients' expectations, and the quality of perioperative supportive care.[16]

The term “complex intervention” can be used to describe a number of multifaceted interventions in healthcare and other domains influencing health and well-being.[5] These include interventions in psychology, mental and behavioural health, nursing, education, public health, social or public policy, among others. The focus of this thesis is on complex interventions in physical domains in healthcare, specifically surgery and rehabilitation, defined here as physiotherapy, occupational and speech-language therapies, and exercise or physical activity interventions.[17,18]

In 2022, the Cochrane Rehabilitation Experts Consortium responded to challenges in defining rehabilitation in the context of research by proposing a definition for rehabilitation that describes it as a “multimodal, person-centred, collaborative process” targeting a person’s capacity and/or contextual factors related to performance with the goal of optimizing persons’ function.[18] The consortium’s definition highlights some of the challenges faced in research in complex interventions in the physical domain, including their multifaceted, highly individualised nature.[18,19] Physical complex interventions can be highly heterogeneous within and between participants, even when they share the same condition or diagnosis.[19] They are also typically tailored or adapted to individual participants or settings, often over time, increasing the scope for variations in their delivery.[3,19,20]

Many complex interventions are also not discrete, rapidly- introduced, “on/off” interventions, and may require sustained interactions between participants and interventionists over long

periods.[19] This can present challenges for monitoring variations in intervention delivery (fidelity) and maintaining participants' adherence during clinical investigations.[3,19–21] Intervention fidelity and participant adherence are distinct but related and interacting concepts. Both concepts are essential for a clinical trial to remain faithful to its protocol. Both can also influence study outcomes, individually or acting together.[3,19–21]

Adherence, sometimes referred to as participant compliance, has been described as participants' acceptance and implementation of prescribed activities in a clinical trial.[22] It represents the extent to which they complied with, performed, were exposed to, or received the intervention in a trial.[23] In pharmaceutical trials, compliance is sometimes used synonymously with "adherence," though adherence is the most frequently used term to describe three main actions by trial participants: initiation of therapy (filling a prescription and taking the first dose), implementation of therapy (correspondence with the prescribed regimen) and persistence with the recommended dosing.[24,25] Adherence in rehabilitation and surgical trials also refers to participants following their randomised group allocation and not crossing over to the comparison condition.[2,4,26,27] In rehabilitation research, adherence is often measured by recording the frequency or number of intervention sessions participants attended, minutes of interventions received, the number of prescribed exercises or activities completed, or the number of times a participant received with an intervention component.[2,4,26,28–31]

Adherence has been investigated extensively, particularly in pharmaceutical research.[24,32,33] Low participant adherence with prescribed treatments is a common problem in pharmaceutical

clinical trials and can considerably reduce the validity and generalizability of their results.[24,32,33] Poor adherence in drug trials can confound safety and efficacy signals, bias estimations of the needed dosage for patient response, increased statistical variability and decreased effect estimates. [24,32,33] For example, several studies had shown the effectiveness of tenofovir disoproxil fumarate (TDF) in pre-exposure prophylaxis (PrEP) to reduce the risk of human immune deficiency virus (HIV) infection in men. However, two studies of PrEP in women in southern Africa did not find evidence for its effectiveness. Analysis of adherence found that only 12% of participants achieved good adherence to the medication throughout their study participation, raising the likelihood that participants in the trial did not take enough of the prescribed dose to see a therapeutic effect.[34]

Adherence in rehabilitation and surgical clinical trials has also been investigated extensively, though the primary focus of research in rehabilitation has been in adherence to treatment in clinical care rather than clinical trials.[19] Several barriers to adherence have been identified, including participant motivation and beliefs, factors linked to the surgical intervention or rehabilitation program, such as discomfort, intervention complexity, time requirements or program length, and practical considerations, such as transportation to intervention sessions, among others.[19,35] All of these factors can reduce participant adherence, moderating their level of exposure to intervention components. This can reduce or mask intervention effects and bias the outcomes of clinical trials.[23] Poor adherence in rehabilitation and surgical trials can also result in decreased statistical power, underestimation of treatment effects, and increase the risk of failing to detect a treatment effect that actually exists, or type II error.[23]

## Intervention fidelity

While participant adherence has been investigated extensively in clinical trials, much less is known about the effects of intervention fidelity in clinical trials of complex interventions.[12] The correspondence between the intervention that was planned and what was actually delivered is the intervention's fidelity.[36–39] Various terms for fidelity are found in complex interventions literature, including treatment, intervention or implementation fidelity, integrity, compliance or adherence.[11,40] Intervention fidelity is, nevertheless, generally conceptualized as the faithfulness of the intervention delivered in a trial to the intervention that was planned in the study protocol.[4–7,13,41–44] Measures of fidelity attempt to monitor and quantify variance in this relationship during clinical trials.[2,15,45–48] Implicit is the idea that intervention fidelity acts as a moderator of the effect of an intervention on outcomes.[49,50]

Several benefits arising from maintaining intervention fidelity in clinical trials of complex interventions have been proposed. High levels of intervention fidelity can reduce random and unintended variability arising from poor intervention delivery, limiting potential confounding from extraneous variables, supporting internal validity[51] and decrease the likelihood of type I and II errors.[37,51–54] Attention to fidelity also decreases the likelihood of Type III error, or a null finding arising from poor quality intervention delivery rather than a null finding arising from intervention ineffectiveness.[55]

Fidelity is also necessary for accurate interpretation of intervention effects.[56] If treatment fidelity is poor or has not been evaluated, additional interventions may be introduced during the trial or contamination of the control group with elements of the experimental intervention may occur.[57] The addition of unplanned, extraneous components or omission of key intervention ingredients can make it difficult to attribute observed effects to the action of the intervention.[16,58] If treatment fidelity is poor or has not been evaluated, one may not be sure that studies' significant results are attributable to the treatment, rather than other, unknown factors, creating in Type I error. If the results are not significant, one can't assume that the poor results are attributable to the treatment rather than addition or omission of other factors, leading to Type II error.[50,52] Correctly rejecting the null hypothesis (no difference between interventions), but for the wrong reason, creates a type III error.[59,60] Poor intervention delivery (fidelity) can lead to nonsignificant outcomes resulting from poor intervention delivery, rather than an actual lack of intervention effectiveness, raising the risk of Type III error.[60,61] The degree of intervention fidelity achieved in a study may be of equal clinical value with quantitative changes when interpreting the results of clinical trials in complex interventions.[62]

## Intervention fidelity and treatment effects

Intervention fidelity may act as a moderator of intervention effects or treatment effect estimates in clinical trials of complex interventions.[39,49] Poor fidelity may lead to variations in intervention delivery resulting in differences in participants' exposure to intervention components (dose).[63] The ProActive trial illustrates how poor intervention fidelity may reduce participants' exposure to the key intervention components, influencing intervention outcomes.[64,65] The ProActive trial

evaluated the effectiveness of a theory-based behavioural intervention for increasing physical activity among sedentary adults at risk of type 2 diabetes.[64] Participants were randomised to one of three trial arms. Participants in the control arm received a brief advice leaflet with information about the benefits of increasing physical activity. Participants in one experimental arm participated in a face-to-face intervention, consisting of one introductory meeting and four in-home sessions delivered at 2, 4, 8 and 16 weeks and two brief phone calls, followed by monthly phone contact up to one year. Participants in a second experimental group received a distance intervention, with the same number of sessions and content delivered by phone and monthly follow-up correspondence.

The interventionists (facilitators ) included a dietician, two nurses and a physical fitness instructor, each of whom participated in a standardised 5-day training course covering the theoretical basis of the intervention, protocol-based intervention delivery, and role-played behaviour change techniques. Interventionists also had access to a comprehensive training manual, protocols for each intervention session, field practice, monthly supervision meetings, bi-weekly team meetings and observation and feedback from the researchers. The researchers also assessed intervention fidelity, monitoring whether the trained facilitators delivered the complex intervention as planned. Facilitators recorded their delivery of techniques after each participant session, and audio recordings of the four in-home intervention sessions were taken then transcribed and coded by the investigators to determine the extent to which intervention components were delivered as planned.

Participants' physical activity levels were measured by self-report with the EPIC-Norfolk Physical Activity Questionnaire a validated physical activity questionnaire[66] and calculation of metabolic equivalents (MET), or the metabolic cost of each activity multiplied by the number of hours participants engaged in the activities per week.[66] At the 1 year assessment, the physical-activity level (Epic-Norfolk scores and METs) of participants who received the intervention, either in person or remotely, did not differ from those who were given a brief advice leaflet.[64] The mean difference in daytime physical-activity level (METs), adjusted for baseline, was -0.04 (95% CI: -0.16, 0.08). Physical activity did not differ between participants in the face-to-face or telephone group (mean difference -0.05; 95% CI: -0.19, 0.10). There was no evidence of a difference between the facilitated theory-based behavioural intervention and the advice leaflet in promoting increased physical activity in an at-risk group.[64]

In-depth fidelity assessment of 108 audio-recorded sessions showed that a median of 44% (inter-quartile range 18-62%) of 14 behaviour-change intervention techniques were applied as per the protocol. Despite the extensive training and ongoing supervision of the interventionists, intervention fidelity was poor, with some intervention sessions providing as little as 18% of the planned techniques. The poor intervention fidelity may have been influenced by a number of factors, including intervention complexity, the quality of intervention tailoring for participants, and the focus of the facilitator training.[65] While the results of the trial may also reflect intervention ineffectiveness, the poor intervention fidelity, reducing participants' exposure (dose) to the key intervention components, may also help explain why there was no evidence of

a difference between the facilitated theory-based behavioural intervention and the brief advice leaflet in promoting increased physical activity in an at-risk group.[67]

In a more recent example, a multicentre, single-blind, parallel, two-arm cluster RCT by Hassiotis et al. (2018) investigated clinical outcomes of staff training in positive behaviour support to reduce challenging behaviour in adults with intellectual disability.[68] Health professionals from community learning disability teams, including speech and language therapists, occupational therapists, nurses and clinical psychologists, were trained to deliver Positive Behaviour Support (PBS), a multicomponent approach using behavioural techniques to reduce challenging behaviours in persons with intellectual and other disabilities.[68] Participating therapists were trained in PBS and study procedures in 2 cohorts over a 15-week period. At the end of training, therapists received a certificate of completion and were paired with one of the trainers as a mentor. Trainers and the investigators also conducted monthly teleconferences and site visits. A training manual supported this process and specified the intervention and behavioural support plan components to be completed when putting the training into practice. The comparison condition was treatment as usual, which included any treatment approach available to community intellectual disability teams within the National Health Service, excluding PBS.

The primary outcome of the trial was challenging behaviour reduction, measured with the Aberrant Behaviour Checklist-Community during participants interviews. Secondary measures included indicators of mental illness, as measured by the Mini Psychopathology Assessment Scale for Adults with Developmental Disability (Mini PASADD).[68] Outcome measures were

completed at baseline, 6, and 12-months. Intervention fidelity was assessed by an independent reviewer, who assessed all the therapists' treatment documentation, including functional assessment, observational data, and the PBS intervention plan. The reviewer also assessed the therapists' intervention plans for their goodness of fit with the PBS using the Behaviour Intervention Plan Quality Evaluation Scoring Guide II. Plans could be classified as weak, underdeveloped, good, or superior.

A total of 215 (87%) and 225 (92%) of participants completed the 6- and 12-month follow-up, respectively.[68] No differences were detected between the intervention and control groups on any of the primary or secondary outcome measures over 12 months. This suggests that there was not a statistically significant different reduction in challenging behaviour in persons with intellectual disability resulting from PBS delivered by community intellectual disability services staff compared to treatment as usual.[68] Although the study had several strengths, including rigorous study design, testing of a single primary outcome, low attrition rate, and an *a priori* analysis plan , poor intervention fidelity was identified as a considerable limitation of the study.[68] Only 30% of participants received all elements of the PBS approach as specified in the training and manual, while a further 43.5% received only one to three of the seven treatment elements, mainly during initial encounters.[68] All of the submitted PBS treatment plans were rated as weak, or not of acceptable quality by the independent reviewer. While other issues, such variations in participant characteristics, therapist workloads, or staff turnover may have influenced the study's findings, poor intervention fidelity was shown to reduce the exposure of participants to key intervention elements.

## Intervention fidelity, statistical power, and treatment effect estimates

Degradation of intervention fidelity may also contribute to biased treatment effect estimates, statistical variance, and reduced statistical power in clinical investigations.[37,50,53,69] Bias refers to a systematic error, or deviation from the truth, in results or inferences."[70] Failure to maintain fidelity to an intervention protocol can increase performance bias in a study.[71] The Agency for Healthcare Research and Quality (AHRQ), defines performance bias as arising from "systematic differences in the care provided to participants and protocol deviation.[71]" Examples of performance bias include contamination of the control group with the exposure or intervention, unbalanced provision of additional interventions or co-interventions, and differences in co-interventions.[71] The AHRQ criteria for risk of performance bias assessment include intervention fidelity (fidelity to protocol), and unintended interventions or co-interventions. The AHRQ also recommends including intervention fidelity as a criterion in the appraisal of individual studies' risk of bias and when assessing their applicability in the body of evidence for research questions.[71]

Poor intervention fidelity may lead to misrepresentation of the true effects associated with an intervention.[72] The direction of the effect of bias arising from poor intervention fidelity on treatment effect estimates has not been established, however. Poor intervention fidelity may either bias treatment effects toward or away from the null by either the addition of extra intervention components or the omission of interventions' active ingredients, accentuating treatment effects, masking them, or reversing their direction.[73–75] The fragility index, a statistical metric for number of events (participant outcomes) required to change a statistically significant

result to a non-significant result,[76] optimal information size, or the minimum amount of information needed to obtain reliable conclusions about an intervention,[77] and other measures have been developed to show the influence of study characteristics on certainty and treatment effect sizes in controlled trials. To the best of our knowledge, no such measure has been derived for intervention fidelity, however.

The impact of study-level characteristics, such as study quality or intervention fidelity, and intervention effect estimates can also be investigated in collections of meta-analyses with meta-epidemiological studies.[78] By contrasting the results of trials with a characteristic of interest with the results of trials without that characteristic, they provide estimates of the average bias associated with that characteristic.[79,80] Poor RCT methodological quality, for example, has been associated with exaggerated (overestimated) treatment effect sizes.[81] Lack of randomization, inadequate sequence generation[82] or randomization concealment,[83] lack of blinding or unclear double-blinding,[78,84] inadequate allocation concealment,[82,85] and lack of prospective trial registration[86] have been associated with significantly increased treatment effect estimates.[78] Other characteristics including selection bias[87], high participant drop out or loss to follow-up,[76] and source of study funding[88] have also been associated with biased RCT treatment effect sizes. The influence of intervention fidelity on treatment effects has been theorized, but not yet shown statistically in meta-analyses of clinical trials in medicine or complex interventions.

Poor intervention fidelity may lead to variability in outcome achievement, inflating error variance in post-test outcomes and decreasing trials' statistical power to detect significant effects and ability to detect differences among treatment arms in clinical investigations.[37,53,63,69,89–92] Even

modest degradations in intervention fidelity may lead to a substantial loss in statistical power resulting from smaller effect sizes, increased error variance, increased variability in outcomes, making it necessary to increase the sample considerably in order to compensate for the loss in fidelity.[53,93,94] In a study with alpha of 0.05 and an intervention with a true population effect size of 0.50, a modest degradation in treatment fidelity (=.75), would lead to a substantial loss in statistical power through making it necessary to increase the sample by fourfold in order to compensate. [53,93]

Reduced participant adherence with trials' treatment protocols can also reduce statistical power in a clinical trial. It has been estimated that poor adherence may limit the statistical power of a study as a square of the proportion of adherent subjects.[92,95] Poor adherence may require sample sizes to be doubled with 30% non-adherence and tripled with 40% non-adherence rates.[30,96] Many complex interventions trials, such as those in physiotherapy, have modest sample sizes, constrained by the availability of resources, participants, and clinical equipoise, and greatly increasing sample sizes may be unfeasible.[90,97] As a result, such studies may be underpowered when intervention fidelity is low or not measured.[53]

Strategies to account for poor participant adherence when interpreting the results of clinical trials include the intention to treat (ITT) principle.[98,99] However, intention-to-treat analyses estimate the effect assignment to an intervention in a trial, not the effect of the treatment itself on participant outcomes.[99] They are agnostic about poor participant adherence, poor

intervention fidelity, protocol deviations such as use of proscribed co-interventions, and anything that happens after randomization.[98,99]

Intention to treat analyses will signal sufficiently large treatment effects (signals), but differences in treatment effects resulting from poor fidelity or adherence may bias their results. The effects of an experimental intervention may be masked or diluted by poor intervention fidelity, reducing ITT analyses sensitivity, or ability to detect a real difference due to treatment.[100]
Per-protocol (PP) analyses, including measures of intervention fidelity, can complement ITT analyses by taking deviations from the study protocol (fidelity) and participants' adherence into account, providing an estimate of the true efficacy of an intervention among those who completed the treatment as planned.[98,99] This topic is explored in greater detail in chapter V.

Our understanding of treatment effects and the appropriateness of clinical trial results to change clinical practice may be further distorted if studies of interventions delivered with poor fidelity drive up statistical heterogeneity in meta-analyses of treatment effects. [2,101] Systematic reviews and meta-analyses of randomized control trials (RCTs) comprise the "gold standard" of evidence for intervention efficacy and often form the basis of clinical practice guidelines and policy decisions.[102,103] However, biases and flaws in the design of RCTs may lead to over or underestimation of the magnitude of treatment effects, skewing the conclusions of meta-analysis when pooled in systematic reviews and meta-analyses.[37,53,69,79,104] Meta-analyses of biased studies may therefore produce unreliable results upon which to inform clinical decision making.

## Defining fidelity

Despite the importance of intervention fidelity for the results, interpretation, and appropriateness to change practice of clinical studies of complex interventions, there is a lack of consensus on its definition or components.[10,11,40,105] Various terms for intervention fidelity, such as fidelity, integrity, adherence, concordance or compliance are found in complex interventions literature. These terms are often ill-defined and used interchangeably or as synonyms for each other to describe concepts related to the accurate delivery or receipt of interventions in clinical trials or practice.[10,11,40,105–108]

For example, while adherence typically refers to whether participants undertake prescribed activities or attend the required number of interventions sessions in a clinical trial, the term "adherence" is also often used when defining fidelity, as in "fidelity refers to adherence to the study protocol."[109] Compliance can refer to either participants' adherence to the required number of intervention sessions[106], or intervention content being delivered in a way that is compliant (faithful) to the intervention manual or maintains treatment integrity.[110] The interchangeable use of terms can lead to confusion and difficulty differentiating between terms such as adherence and fidelity.

Three of the most common terms are fidelity, integrity, and adherence.[11,40] The three terms have similar and sometimes overlapping etymological roots and their current definitions often include at least one of the other two words. Shared among these terms are notions of similarity, accuracy, or faithfulness that characterise intervention fidelity. The word *fidelity* entered the English language in the XVth century *as fidelite*, from Middle French *fidelité*, derived from

the Latin word *fidēlis* ( *fidēlitāt-, fidēlitās, fidelitatem),* meaning "faithful, loyal, or trustworthy.[111]" For centuries, *fidelity* has also been used to refer to accuracy, as in "The movie's director insisted on total fidelity to the book.[111]" Fidelity is currently defined as the quality or state of being faithful, accuracy in details or exactness, and the degree to which the detail and quality of an original, such as a picture, sound, or story, is copied exactly.[111]

Two other terms often used interchangeably with fidelity to describe intervention delivery, integrity, and adherence, share similar etymologies. Integrity entered the English language at the beginning of the XVth century as *integrite* ("wholeness, perfect condition") through Old French *integrité* ("innocence, blamelessness; chastity, purity,"), derived directly from Latin *integritatem* (*integritas*, ***integer***), or "soundness, wholeness, completeness, purity, correctness, blamelessness.[112]" Today, *integrity* is defined as firm, complete or whole adherence to a code.[111] Meanwhile, *adherence* entered the English language in the XVth century from Old French *adherence (*steady attachment of the mind or feelings to a person, cause, belief, etc.), derived from Latin *adhaerentia, adhaerent, adhaerare* ("to stick to").[112] It is currently defined as steady or faithful attachment, or fidelity.[111]

Other terms used to describe intervention fidelity also entered English through Old French from Latin and express ideas , such as *concordance,* or "a state in which things agree and do not conflict with each other[111]," from Old French *concordance* ( "agreement, harmony"), derived from Latin *concordare* "be of one mind, or of the same mind").[111,112] Compliance, defined as "the act of conforming, submitting, or adapting (as to a regulation or to another's wishes) as required

or requested[111]," entered English from Old French *complir* ("to accomplish, fulfil, carry out"), derived from Latin *complere* "to carry out or fill up.[111,112]"

The terms "compliance" and "adherence" in particular are often used interchangeably in healthcare and fidelity literature to denote patients' or participants' behaviour, but carry different and evolving connotations.[107,108] Sackett and Haynes first defined "compliance" in relation to prescribed medical regimens in the late 1970s as "the extent to which the patients' behaviour (in terms of taking medications, following diets, or following lifestyle changes) coincide with medical advice[113]." With a shift towards patient-centred care, however, compliance took on paternalistic connotations, implying patient disobedience, lack of patient autonomy, and limited patient involvement in decision-making.[108,113,114] By the 1990s, a shift towards the term adherence to define the degree to which patients followed a recommended therapeutic regimen took place.[108] The World Health Organization (WHO)'s Adherence Project defined adherence as "the extent which a person's behaviour-taking medications, following a diet, and/or executing lifestyle changes, corresponds with agreed recommendations from the health care provider.[115]"

## Intervention fidelity: A rose by many other names

A plethora of other terms, definitions, and descriptors for intervention fidelity can be found in complex interventions literature (see chapter II). These vary between disciplines – for instance between Education, Public Health, Physiotherapy, and Behavioural Medicine, and context.[10,11,40,105] Among these, intervention fidelity has been conceptualized as the faithfulness

of an intervention to its underlying theoretical and therapeutic principles or to clinical guidelines[42–44], intervention delivery in line with a plan[41] or study protocol[41–44], the concordance of an intervention, as delivered, to the intervention as planned, and others. Although these descriptions vary in in content and detail, they are conceptually similar and describe "the extent to which an intervention was delivered as conceived and planned.[116]" Implicit among them is the idea that intervention fidelity may act as a moderator of the effect of an intervention on outcomes[49], and necessary to arrive at valid conclusions concerning interventions' effectiveness in achieving the target outcomes. [10,11,40,105,116]

The lack of consistency in the terms and definitions for intervention fidelity can be elucidated by tracing its development from its origins in psychotherapy to its adoption across a range of complex interventions, research disciplines, and settings. Overlapping terms and a lack of consensus about the components of intervention fidelity can be seen through its evolution. In surgery and rehabilitation, interest in intervention fidelity has been more recent, and grown consistently over the past decade. However, the adoption of fidelity models and constructs from a variety other fields and lack of consensus on elements of fidelity most suitable for rehabilitation interventions mean that those who develop, implement and study physical complex interventions have no common language by which they can make assessments and develop robust methods to support intervention fidelity. [10,40,63,117]

## The history and development of intervention fidelity

The concepts behind intervention fidelity emerged from psychology and psychotherapy in the 1950s and 1960s. Over the following 70 years, its definition and components continued to evolve, broadened to include a range of methodological techniques employed to monitor and enhance the reliability and validity of interventions across a wide range of clinical and research disciplines.[39] This expansion has resulted in the many different definitions and fidelity constructs used inconsistently and interchangeably today. Interest in intervention fidelity has grown in surgery and rehabilitation in the past 2 decades, but its development in physical complex interventions is more recent than in psychology and behavioural health, and fewer publications in these fields describe the development of logic models for intervention fidelity in their clinical trials or monitoring of fidelity with objective measures.[36,46,52,60,105,118,119]

### 1950s-1980s: Fidelity evolves in psychotherapy, psychology, and behavioural health

As early as the late 1950s, concerns arose about overlapping treatment conditions, vague intervention descriptions and dosage, and poor reporting hindering replication.[120] The evolution of client-centred psychotherapy increased the systematic evaluation of interventions and their implementation in psychology.[121] Intervention fidelity and fidelity measurement were first delineated in Behavioural Psychology and Psychotherapy literature in the late 1970s.[122] Emphasis was placed on treatment integrity and collecting data during treatment sessions to ensure that intervention procedures followed protocols, treatment differentiation and therapist competence were maintained, and treatment length was held constant.[123] By 1979, Cook and Campbell proposed that absent fidelity monitoring or optimization would make it difficult to

know if significant results in trials of new interventions (in psychology) are due to an effective treatment or to unknown factors that may have been unintentionally added to or omitted from it during the study.[124]

Still referred to as treatment integrity (or treatment adherence) in psychology literature by the early 1980s, it was increasingly defined as the degree to which a treatment is implemented as intended.[125,126] Few studies in psychotherapy or behavioural psychology monitored or reported fidelity, however, and poor intervention replicability continued to concern researchers and research methodologists. In 1980, Billingsley, White, and Munson (1980) reviewed 108 studies of applied behavioural analysis and behaviour modification and found that while 82% of studies reported the reliability of their outcome measures, only 5.6% of those same studies assessed treatment implementation (fidelity).[127] In 1982, Peterson et al. found that only 20% of 539 behavioural psychology studies published between 1968 to 1980 assessed treatment implementation.[126]

### 1980s-1990s: Fidelity in community mental health, education, school-based interventions

By the 1980s, there was growing interest in identifying the essential features of interventions and enhancing their implementation in other complex interventions. It was increasingly applied to interventions in community mental health, education, and school-based intervention programs. Blakely et al. (1987) described social intervention programs as consisting of a finite number of components and defined fidelity as the proportion of program components that were actually implemented.[128] In 1988, Breck and Test empirically measured program

implementation, including fidelity to their conceptual models, in community mental health settings, demonstrating the feasibility of systematic fidelity monitoring.[129] Their research and series of publications anticipated more systematic fidelity monitoring and the development of fidelity assessment tools. Increasing attention was also being paid to intervention fidelity and fidelity monitoring in education and school-based interventions research.[130] Meyer and Herman also examined the implementation of school-based programs, recommending that modifying interventions to accommodate local needs is necessary and acceptable as long as the critical features of a program are delivered as planned.[131]

### Healthcare, health service delivery, medical care

In the 1980s and 1990s, greater recognition the effects of intervention fidelity on the internal validity of intervention studies and its importance for both quality healthcare and the dissemination of evidence-based interventions led to increasing efforts to formalise and guide fidelity assessment to enhance intervention fidelity.[132,133] Donabedian's model (1982) for evaluating the quality of medical care proposed that the quality of care includes aspects of both structure, encompassing the framework for medical service or interventions delivery, and process , comprising the way in which the services or interventions are delivered.[134] Fidelity constructs were also expanding to include additional dimensions beyond treatment integrity and differentiation.

### 1990- 2004: The first intervention fidelity guidelines

Moncher and Prinz proposed the first set of guidelines for enhancing of treatment fidelity in

1991.[37] Moncher and Prinz recommended that researchers operationally define interventions, adequately train interventionists using treatment manuals and provide them with ongoing supervision, measure adherence to treatment (fidelity) through observation, and use fidelity data to interpret research findings.[37] Subsequently, Lichstein, Riedel, and Grieve's treatment implementation model (1994) proposed two additional processes as part of intervention fidelity: treatment receipt, or the degree to which participants understand and demonstrates knowledge of and ability to use treatment skills, and treatment enactment, or the degree to which the participants apply skills learned in treatment in their daily lives.[135] Lichstein and colleagues also identified strategies to enhance treatment fidelity, a number of potential threats to fidelity, and methods of measuring treatment fidelity in research.[135] Their work, along with Donabedian's model and Moncher and Prinz' recommendations influenced future development of intervention fidelity in research and intervention implementation in clinical practice, and continue to be key components of fidelity enhancement to this day.

Throughout the 1990s, interest in refining fidelity assessment tools continued to grow with increasing recognition that without data from fidelity monitoring, it would be "difficult to determine whether non-significant results are due to a poorly conceptualized program or to an inadequate or incomplete delivery of the prescribed services.[136] Moncher and Prinz (1991) had defined fidelity as "confirmation that the manipulation of the independent variable occurred as planned," describing two principle components to treatment fidelity. The first, treatment integrity, is the degree to which a treatment condition is implemented as intended. The other, treatment differentiation, "refers to whether treatment conditions differ from one another in

the intended manner such that the manipulation of the independent variable actually occurred as planned.[37]" In 1994, McGrew et al. were among the first to show a predictive relationship between these factors and treatment effectiveness.[137] They worked with an expert panel to identify the critical ingredients of an evidence-based practice in community mental health (Assertive Community Treatment -ACT) and construct a fidelity assessment scale and fidelity index. They then evaluated a collection of published studies of ACT with their fidelity scale, demonstrating a predictive relationship between their fidelity index and client outcomes.[137]

### Continued development in psychology and expansion in education and public health

Intervention fidelity monitoring continued to develop within psychology, education, and public health. By 1998, Dane and Schneider recommended assessment and verification of intervention fidelity (integrity) as a vital component of intervention evaluation.[136] They expanded on the components of intervention fidelity, describing it in five dimensions: 1. adherence, or the extent to which critical components are implemented as intended (also referring to it as treatment integrity); 2. quality as a measure of instructional quality separate from adherence to components; 3. exposure, or the amount of instruction provided (dosage); 4. participant responsiveness, or the extent to which participants responded to the intended treatment (treatment receipt); and 5. program differentiation, or the extent to which a treatment differs from the comparison condition.[136]

Building on Moncher and Prinz[37] and Dane and Schneider's[136] work, the definition, monitoring and measurement of intervention fidelity expanded in the health sciences. A consortium of

National Institutes of Health (NIH) -funded projects focused on improving health-related outcomes culminated in the formation of the NIH Treatment Fidelity Workgroup in 1999. The fidelity working group made several recommendations to improve treatment fidelity in health research, publishing the NIH Behaviour Change Consortium Comprehensive Fidelity Framework (NIH-BCC) in 2004.[133] The framework provided guidance for the assessment, enhancement and monitoring of intervention fidelity in health behaviour research.[39,133]

### The NIH Behaviour Change Consortium Comprehensive Fidelity Framework

The NIH-BCC framework conceptualised fidelity across five core domains: Study Design, Provider Training, Intervention Delivery, Intervention Receipt and Enactment (table 1.1,).[133] The NIH-BCC fidelity framework recommended actions to help researchers enhance and monitor intervention fidelity.[39,133] These include: 1. training and supervising intervention providers using specific strategies, and assess the providers before treatment delivery to ensure they acquired critical skills; 2. measuring treatment adherence and dosage, but also variations in treatment fidelity among interventionists, treatment differentiation, and treatment receipt; 3. collecting data on treatment and comparison sessions using audio tapes or observations and conducting exit interviews with comparison group interventionists to assess treatment differentiation; and 4. monitoring and describing treatment receipt to ensure participants used the learned intervention skills in day to day settings. The NIH-BCC recommendations were later adapted for public health trials.[138]

| | | |
|---|---|---|
| **NIHBCC Comprehensive Fidelity Framework,** Borelli et al, 2005 | Study Design | Does study adequately test its hypotheses in relation to underlying theoretical and clinical processes? Are interventions' active ingredients fully operationalized? |
| | Provider Training | Is there standardizing training between providers, ensuring providers are trained to criterion, and is there monitoring and maintaining of provider skills over time? |
| | Treatment Delivery | Is there differentiation (providers only deliver the target treatment and not other treatments), treatment competency (providers maintain skills learned in training) and treatment adherence (delivery of treatment components as intended)? |
| | Treatment Receipt | Was the treatment delivered to the participant actually "received," or understood accurately by participants? Did participants demonstrate knowledge of, and ability to use, the skills or recommendations learned in treatment. |
| | Treatment Enactment | Can participants perform treatment related behavioral skills and cognitive strategies in relevant real-life settings? Are skills implemented in appropriate situations and time to have the intended effect on clinical and research outcomes? |
| **Fidelity Framework** by Gearing et al, | Intervention design | Intervention's framework and elements essential to the design of a trial, it's evaluation or replication. Includes a program model, treatment manual. |
| | Intervention training | Intervention fidelity requires adequate training and supervision of interventionists. Training elements include interventionist differences, such as levels of skills, education, experience, and implementation styles. |
| | Intervention delivery | Adherence to and integrity of treatment. Measures may include; frequency counts, logs, records of how many times a behavior or technique occur. Includes whether prescribed behaviors have taken place, dose of program content delivered to and received by participant. |
| | Intervention receipt | Elements that focus on whether participants received the treatment, and whether essential elements were provided in the treatment. |

**Table 1.1:** The NIH-BCC Fidelity Framework and Gearing's Comprehensive Fidelity Framework

## 2007-present day: Further development of fidelity models and assessment

Further development of intervention fidelity constructs and models in health and behavioural sciences continued to expand the dimensions of fidelity, with the Implementation Fidelity Framework (IFF) by Carroll et al. in 2007[2], Gearing's Comprehensive Fidelity Framework[40] in 2011, and Nelson et al.'s 5-step framework assessing intervention fidelity in experiments in educational and behavioural interventions in 2012. The IFF conceptualises fidelity in terms of intervention or program content (active ingredients), coverage (the reach of the intervention), frequency (the number of sessions), duration (time taken), complexity (difficulty level of the intervention or program), quality of delivery (interventionist competence), facilitation strategies (strategies used to enhance intervention delivery) and participant responsiveness (participant

receipt).[2] The IFF also added an emphasis on the context interventions are delivered in and moderating factors that may influence the intervention and participants' responses to it. Gearing's Comprehensive Fidelity Framework (table 1. 1), described in greater detail in Chapter II of this thesis (see Chapter II, table 2.3), operationalised intervention fidelity in four elements: design, provider training, monitoring of intervention delivery, and intervention receipt, with special consideration given to threats to fidelity and their measurement.[40]

Nelson et al. defined fidelity as "the extent to which an intervention's core components have been implemented (and differentiated from control conditions) as planned.[139]" In Nelson's model, fidelity assessment is the measurement of the causes (interventions) of effects (outcomes). Intervention fidelity calls for measuring all causes, to the extent possible, in order to understand what works in the context of an RCT.[139] Assessing intervention fidelity provides a means for describing the extent to which the intervention as implemented (the actual cause) resembles the intervention as designed (the theoretical cause), providing an explanation of how assignment to a condition (experimental or control) resulted in some difference (or lack of difference) between the conditions.

Mirroring aspects of the NIH-BCC fidelity framework, Nelson's five-step process for assessing intervention recommends:

1. Specify the intervention model, or the theoretical mechanisms by which the intervention is meant to work and effect change in an outcome.

2. Identify appropriate fidelity indices, or what will be assessed to indicate intervention fidelity and how it will be measured.

3. Determine the reliability and validity of the fidelity index.

4. Combine indices where appropriate (if multiple fidelity assessments or assessment of multiple intervention components is undertaken).

5. Link fidelity to outcomes where possible. Linking analyses linking fidelity measures to outcome measures, revealing how differences in fidelity may be associated with differences in outcomes

## Emergence of Implementation Science

The number of fidelity models and constructs continued to grow over the past two decades and a new scientific discipline, Implementation Science emerged. Implementation Science focuses on facilitating the spread and uptake of evidence-based practices in healthcare and public health, and is defined by the NIH as "the study of methods to promote the adoption and integration of evidence-based practices, interventions, and policies into routine health care and public health settings to improve the impact on population health. [140,141]" Implementation science also examines the fidelity with which practitioners and policymakers implement evidence-based practice and research into regular and adaptation of these to local contexts. The focus of this thesis is intervention fidelity in clinical trials, whether of intervention efficacy or effectiveness, however and not on the implementation of interventions in clinical practice. Therefore, the development and current state of evidence for implementation science or fidelity models and assessments used therein are not captured by this thesis.

## Proliferation of fidelity constructs: Lack of consensus on defining or operationalising fidelity

The proliferation of fidelity models, frameworks, and related measures, both for use in studies of interventions or program effectiveness and the implementation of evidence-based intervention in clinical practice, continued throughout the past two decades. For example, Gearing's 2011 review of peer-reviewed literature reviews, systemic reviews, and meta-analytic review articles in health and behavioural health published between 1190 and 2009 identified 27 fidelity models or constructs applied in empirical papers in psychology or behavioural health. By 2017, Watson et al.'s systematic review of intervention fidelity in physical activity interventions identified 66 approaches to measuring fidelity across 21 studies , with 52 approaches measuring intervention delivery fidelity, eight measuring enactment, four measuring receipt, two measuring training fidelity and no papers assessing fidelity in intervention or studies' design.

Fidelity models, frameworks, constructs and assessments can now be found in a range of fields, including psychology and mental health[132], behavioural health[142], education[139], physiotherapy[60], occupational therapy[143], speech-language pathology[144], nursing[145], public health[146], and others. As fidelity evolved from its initial focus on treatment integrity into a broader multi-faceted concept, it encompassed an increasing number of elements. Ultimately, the concept grew to include the fidelity of each stage of an intervention where an action is intended to occur, as well as the design of the research study in which they take place.[11] The multifaceted nature of intervention fidelity, together with the absence of a unified approach to enhancing or monitoring it within and across research disciplines (e.g., education, health, social programs),

have led to a broad range of terms used to describe it (i.e. fidelity, integrity, adherence, competence, compliance, concordance), and the varied operational definitions for fidelity and methods to assess it.

## Consensus statements: Fidelity in reporting guidelines

With growing emphasis on improving the reporting and replication of interventions from clinical trials, research reporting guidelines such as the Transparent Reporting of Evaluations with Nonrandomized Designs (TREND) of 2004[147] and updates to research reporting guidelines Consolidated Standards of Reporting Trials (CONSORT)[148] guidelines in 2010 and Consolidated Standards of Reporting in Non-pharmacological treatments (CONSORT-NPT)[149] in 2017 included items related to intervention fidelity or participant adherence to study protocols. For example, TREND item 4 includes detailed description of the intervention intended and how they were actually delivered, including their dosage, providers, and any strategies used to increase "compliance or adherence.[147]" The CONSORT[148] statement also recognises that a trial protocol may not be followed fully by some trial participants for a wide variety of reasons, including failure to receive the entire intervention as planned or received a proscribed co-intervention.[103] The CONSORT-NPT includes 2 items addressing intervention fidelity: item 5c describes "details of whether and how adherence of care providers to the protocol was assessed or enhanced," while 5d describes this for participants' adherence to interventions.[149]

The Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT)[26] statement, published in 2013, Template for Intervention Description and Replication (TIDieR) Checklist[150] in

2014 , and the Consensus on Exercise Reporting Template (CERT)[151] of 2016 also included items reporting intervention fidelity and participant adherence (appendix I figures 1.1-1.3). The SPIRIT[26] statement includes one item (11b), describing criteria for intervention modifications for a given trial participant, though it offers little guidance on how those modifications can be described or identified in study protocols, and 11c, strategies to improve adherence to intervention protocols, and any procedures for monitoring adherence. Both TIDieR (9-12)[150] and CERT (14-16)[151] included items relevant for reporting intervention adaptation and fidelity.

The TIDieR[150] checklist includes 12 items, with items 8 "When and how much" (describes the dose of an intervention), item 9 "Tailoring" (if the intervention is planned to be personalized, titrated or adapted, then describe what, why, when, and how) item 10 "Modification" (if modified during delivery, what, why, when, and how), and items 11 and 12 "How well (intervention fidelity and adherence and assessment) representing key aspects of intervention specification, delivery and fidelity. The CERT[151] checklist includes item 5 (detailed description of how adherence to exercise is measured and reported), items 13 (when and how much, or dosage), items 14a and 14b (tailoring and adaption, including detailed description of how exercises are tailored to the individual) and items 16a and 16b (describe how adherence or fidelity were measured, and extent to which the intervention was delivered as planned).

### Where are we now? Fidelity in education, psychology, behavioural and public health

While it is increasingly recognized that intervention fidelity is important for the interpretation and replication of clinical trials in complex interventions, few studies measure or report

intervention fidelity, and fewer measure fidelity quantitatively.[1,50,152] Systematic reviews of intervention fidelity in education, psychology, behavioural and public health have identified poor use and monitoring of intervention fidelity.[11,139,153–155] For example, Walton et al. found that fewer than just over a third of the 66 studies included in their systematic review of complex, face-to-face health behaviour change interventions measured intervention fidelity.[142] A systematic review of school-based reading interventions in education by Capin et al. (2018) also found that less than half (47%) of the 175 included studies reported treatment fidelity data, either numerically or in narrative form.[156] Kechter et al. (2019) reviewed reporting of intervention fidelity in mindfulness-based intervention trials, finding that only 25 (12%) of 202 eligible papers described or reported intervention fidelity.[157] Other reviews have also found poor identification of fidelity constructs or components. A scoping review of fidelity by Slaughter et al. (2015) also review found that none of 72 included papers reported either fidelity definitions or conceptual fidelity models or frameworks.[10,158] Lambert et al. (2017) also found a lack of attention to the quality of fidelity assessments, when they were reported, with few studies using objective methods.[158] Factors that may influence the poor uptake or reporting of intervention fidelity are discussed later in this chapter.

### The ”physical” complex interventions: Surgery, rehabilitation, physical activity interventions.

In parallel to the development of intervention fidelity in the 1990s and 2000s, growing interest was focused on complex interventions in healthcare. The focus of this thesis is on complex interventions in controlled trials in physical domains of health care, specifically surgery, physiotherapy, occupational therapy, speech and language therapy (rehabilitation, broadly

defined), exercise and physical activity interventions. These practitioner-based complex interventions share key characteristics, including involving multiple interacting components targeting a range of effects to achieve a desired outcome, requiring a minimum amount of practitioner skill and expertise to deliver the intervention, practitioner learning curves, and some necessary degree of adaptability or flexibility to be tailored to individual participants, interventionists, or settings. These characteristics present unique challenges and opportunities for their innovation and evidence-gathering, such as the need for replicable intervention implementation despite the inherent difficulty in standardizing interventions which are multifaceted and often tailored to individuals.[3–5] These complexities must be considered in their study design, in addition to the practical and methodological difficulties associated with most interventions.[5–7]

## The Medical Research Council Framework for developing and evaluating complex interventions

The United Kingdom's Medical Research Council (MRC) published its first framework for developing and evaluating complex interventions in in health and social care services, public health practice, and other areas of social and economic policy 2000.[159] This was revised in 2008[5] and most recently, in 2021[20]. The MRC framework proposes that complex intervention research goes beyond focusing mostly on intervention efficacy or effectiveness, but also asks a broader range of questions, including why and how an intervention works, and it interacts with the context in which it is implemented, among others (figure 1.1). Interventions are considered to be complex because they are made up of various interconnecting parts, target a range of

behaviours, require expertise and skills by those delivering and receiving the intervention; target a number of groups or settings, and permit a level of implementation flexibility.[20] Their evaluation is made difficult by the complexities of developing, identifying, documenting, and reproducing the intervention and the range of settings and contexts in which they may be implemented.[5,159] The development and evaluations of complex interventions requires a phased, often non-linear approach and use of qualitative and quantitative evidence. Preliminary work is often required to establish the active components, parameters, and feasibility of the intervention so that they can be delivered effectively and evaluated in a trial.

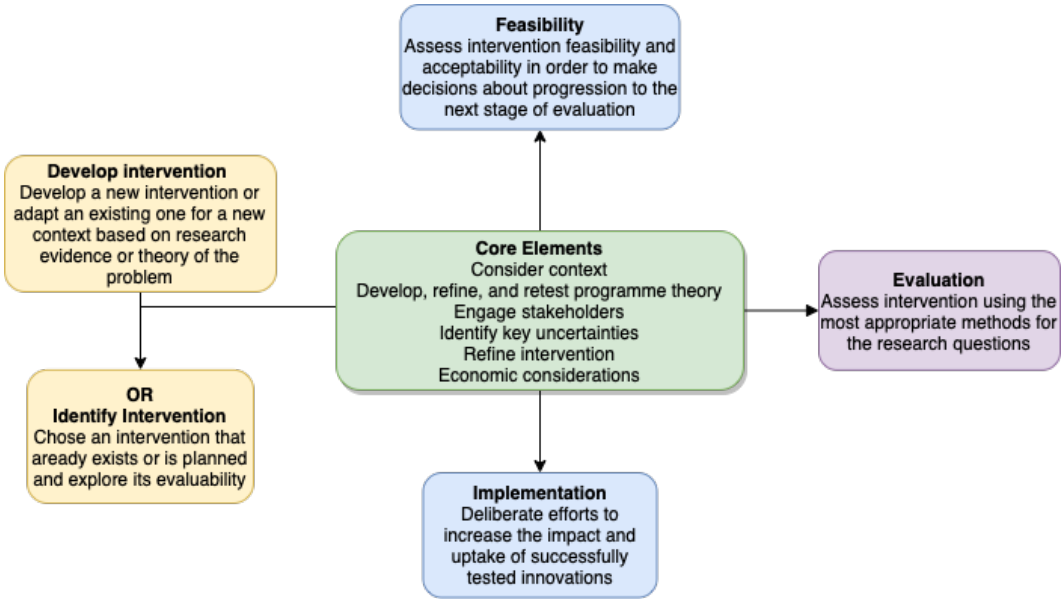



**Figure 1.1:** MRC Complex interventions guidance

Phases and core elements of complex intervention research. Adapted from Skivington et al, 2021[20]

The MRC framework recognises that many complex interventions may require tailoring to local circumstances rather than completely standardised delivery or implementation. As a result, intervention fidelity for complex interventions may not be straightforward and may depend on

the purpose and context of the study investigating them. In an explanatory study or one seeking to identify interventions' active ingredients, greater standardization and fidelity of intervention delivery would be required. In pragmatic trials or interventions designed to be adapted to local circumstances, greater adaptation and variations in intervention implementation may be permissible.

The 2008 update of the MRC guidelines recognised the value of process evaluation within trials, including that they: "assess fidelity and quality of implementation, clarify causal mechanisms and identify contextual factors associated with variation in outcomes.[49]" In 2015, the MRC process evaluation guidance was published, emphasizing the importance of intervention fidelity for understanding intervention effects.[49] Intervention fidelity was defined as whether an intervention was delivered as intended, with dose (the quantity of intervention implemented) presented as an independent concept. Findings for limited intervention effects in a study could reflect weaknesses in intervention design or poor intervention implementation, though positive intervention outcomes could be found even when interventions were not delivered as intended, highlighting the importance of fidelity assessment in process evaluations to help determine what was done, what was changed, and what may (or may not) have worked.[49]

Referencing intervention tailoring for different contexts to interventions' theoretical underpinnings is also recommended to differentiate adaptations facilitating intervention use in new contexts from those that undermine intervention fidelity. Of note, the 2015 process evaluation guidance discusses the lack of consensus on how best to divide intervention implementation into components, such as fidelity, dose, and reach, and the impossibility of

adjudicating between the various implementation and fidelity frameworks currently available. [49]

## The IDEAL Collaboration and IDEAL Frameworks

Between 2007 and 2009, the Balliol Collaboration, an international group of experts in surgery, evidence-based medicine, and clinical trials methodology took part in three conferences on surgical innovation and evaluation at the University of Oxford. Their primary goal was to draft a series of papers for The Lancet describing the relationship between innovation and clinical practice in surgery. The first paper focused on the process of innovation and assessment of novel surgical interventions.[160] The second paper described the challenges faced in designing clinical trials of surgical interventions[161], while the third paper presented the IDEAL (Idea, Development, Exploration, Long-term study) framework to aid the "timely and appropriate assessment of surgical innovation along its different stages.[3]"

The IDEAL framework was later expanded to include other non-pharmaceutical complex interventions such as surgically implanted medical devices IDEAL-D[162] in 2016), radiotherapy (R-IDEAL[163] in 2017), and most recently, physiotherapy and rehabilitation (IDEAL-Physio) in 2018.[7] The framework was updated with new recommendations and an additional stage, IDEAL-0, for preclinical studies and development in medical devices in 2019.[164] The IDEAL reporting guidelines were published in 2021.[165] As the focus of this thesis is in complex interventions in the physical domain (surgery and rehabilitation), the content and recommendations of the main IDEAL framework and IDEAL-Physio (rehabilitation) will be discussed. Subtle variations between the extensions of the framework exist, given their target interventions (for example, IDEAL-D[162]

focuses on innovation in implantable medical devices).

The IDEAL framework was based on the principle that innovation and evaluation in clinical practice can, and should, evolve together in an ordered manner: from conception to development and then to validation by appropriate clinical studies and, finally, longer-term follow-up.[3,7] The IDEAL framework divides the stages of non-pharmacological complex interventions' innovation into 5 phases: 1. idea; 2a. development; 2b. exploration; 3. assessment and 4. long-term study (figure 1.2). In stage 1, the idea phase, proof of concept or first-in-human use of the intervention takes place and formal data collection should begin. Stage 2a is the phase for iterative improvement and adjustment of the intervention to achieve stability and replicability, with thorough data recording. Stage 2b involves achieving consensus on the intervention's stability through formal evaluation using systematically collected group or cohort data. Stage 3 is the phase for formal comparative assessment of treatment, usually involving randomized studies. Stage 4 involves long-term follow-up.

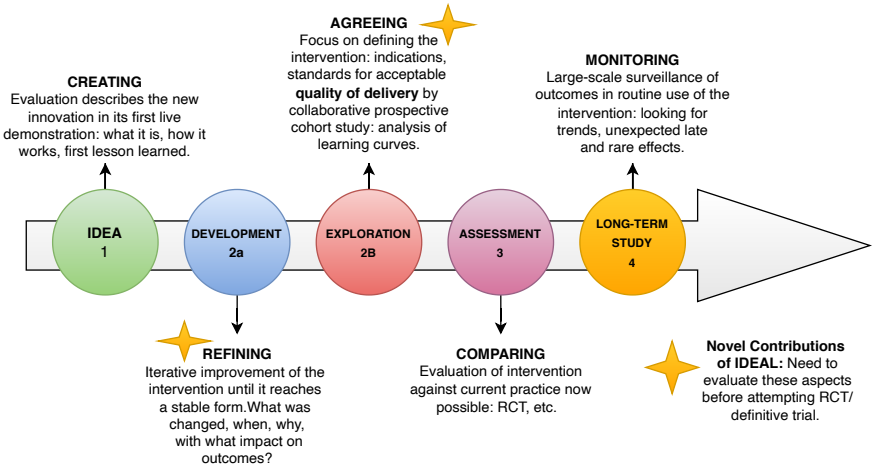



**Figure 1.2:** The IDEAL Framework: Stages and evaluation pathway.

While terms for intervention fidelity are not explicitly included in the IDEAL framework's recommendations, IDEAL recommends actions that are essential for intervention fidelity throughout the stages leading up to, and including, controlled trials. These include detailed documentation and testing of intervention procedures and parameters, documentation of iterative modifications to interventions and reasons for the modifications, assessment and consideration of interventionist learning curves with new procedures, documentation intervention tailoring and adaptation, with reasons for the tailoring and its outcomes, and ensuring interventions are sufficiently stable and replicable before being subjected to comparative evaluation in controlled trials.[7,164]

IDEAL Stage 1 (IDEA: first in human or proof of concept studies), recommends careful documentation of the evolving intervention, including the interventions' theory and proposed mechanisms of action, intervention details and any modifications that were made to the it, with reasons.[7,164] It also recommends documenting interventionists and patients (or participants') experience with the intervention. Use of a structured reporting system, such as the SCARE guidelines is also recommended. The Surgical Case Report Guidelines (SCARE[166,167]) include items for detailed intervention description (including physiotherapy), procedures, and participant adherence.

Stage 2a (Exploration: Single centre/single intervention; case series and prospective cohort studies) recommends detailed and technical description of the intervention and its procedures with careful documentation of modifications made to the intervention with reasons for all

changes. [7,164] Outcomes at this stage are mainly safety and technical and procedural stability. By the end of stage 2a, the intervention should have detailed procedures, defined parameters and indication, with no further iterative modifications expected. The intervention should be sufficiently stable to allow for high quality replication by a larger number of interventionists and centres in the following stage, IDEAL 2b. The study protocols for IDEAL stages 2a and 2b should also be designed with these considerations in mind and be registered and available for consultation and scrutiny.

IDEAL stage 2b (Exploration) acts as a bridge from observational to comparative evaluation, with prospective multicentre exploratory studies and pilot or feasibility trials. [7,164] Studies in this stage tend to be collaborative and multicentre, with cooperative data collection. The focus of this stage includes intervention and trial feasibility and determining whether an intervention is sufficiently stable and defined that can be more widely adopted and a progression towards a comparative trial (RCT) is feasible and desirable. Key features supporting intervention fidelity in this stage include standardizing intervention and comparator definitions, defining and measuring intervention quality standards, continued analysis of interventionist learning curves and training needs, and estimating the influence of prespecified tailoring and adaption on participant outcomes, analysis of any unexcepted or adverse events, and collecting data for stakeholder values and experiences with the intervention.

In state 3 (Assessment), comparative effectiveness testing of the intervention takes place, typically though an RCT or alternative designs (i.e. cluster, preference RCTs, stepped wedge or

adaptive designs). [7,164] The two main endpoints of this state are valid evidence for the relative effectiveness of the intervention and identification of issues requiring long term monitoring, if applicable. Key recommendations include use of Core Outcome Measures in Effectiveness Trials (COMET)[168], those made by the 2010 CONSORT[148] update with extension for non-pharmacological treatments, and use of appropriate reporting guidelines such as SPIRIT[26] for protocols, and TIDieR[150], both of which include items reporting intervention fidelity and participant adherence. Stage 4 (Long-term monitoring) is a surveillance stage and takes place mostly through mentoring of registry or routine care databases and lies beyond the focus of intervention fidelity.

## Intervention Fidelity in surgical RCTs

In 2015, Blencowe et al., working with the MRC Hub for Trials Methodology Research Network Workshop, published further guidance for interventions in randomised trials in surgery, including pre and post-operative physiotherapy.[16] They identified key issues to consider in the design of surgical RCTs, including identification of each surgical intervention and their components, who will deliver the interventions, and where and how the interventions will be standardised and monitored during the trial. They defined intervention fidelity as "How far those responsible for delivering an intervention actually adhere to the intervention as it is outlined by its designers.[2] Fidelity is also referred to as compliance or adherence.

Guided by Carroll's Implementation Fidelity Framework, Blencowe et al. proposed eight questions to ask when designing an RCT for a surgical intervention (table 1.2). These encompass

definition of the intervention and its comparison, pre-determining how interventions will be specified in the RCT, who will deliver the interventions, and how their delivery will be monitored during the trial to ensure intervention fidelity. As with the IDEAL Framework's recommendations, Blencowe et al. also recommend the skills of all these team members may influence treatment outcomes, and it may be necessary to account for them in RCTs evaluating surgical interventions, both at the design and analysis stage.

| | Questions to ask during the design of surgical interventions randomised controlled trials (RCTs) |
|---|---|
| 1. | Does the RCT involve a surgical intervention? |
| 2. | What is/are the surgical intervention(s) under evaluation? |
| 3. | What is/are the concomitant intervention(s) accompanying surgery? |
| 4. | What will influence standardisation of the interventions?<br>a. What is the overall study design?<br>b. What type of comparator is in the RCT?<br>c. In what stage of development is/are the surgical intervention(s)? |
| 5. | How will the intervention(s) be standardised in the RCT? |
| 6. | How will delivery of the intervention(s) be monitored in the RCT (fidelity)? |
| 7. | Who will deliver the intervention(s) (operator expertise)? |
| 8. | Where will the intervention(s) be delivered (context)? |

**Table 1.2:** Blencowe et al.: Questions to ask during the design of surgical interventions RCTs[16]

The recommendations emphasized the importance of pre-specifying how standardized the surgical intervention should be in the trial protocol. This would include identifying which components or steps of a surgical operation are essential *a priori*, so that these, and any prohibited steps performed in error or by necessity can be monitored during the trial. Assessing whether key components of an intervention were performed according to protocol could be

accomplished by direct observation of the intervention, video or photographic assessment, or surgeon-reported descriptions of what was done. In trials with a more flexible protocol, stating that these intervention steps can be carried out according to surgeons' preferred techniques, monitoring intervention delivery may not be required, particularly. This loss of information about intervention delivery and fidelity may limit interpretation of trial results, however. In explanatory trials, intervention fidelity data could be used to improve training of surgeons if fidelity is poor.

Subsequent work by Blencowe et al. developed a typology and framework to standardize the delivery and monitoring of surgical interventions in their trials and protocols.[169] Included in this typology , intervention fidelity is essential for the interpretation of surgical trial results and subsequent implementation of the interventions in clinical practice. Intervention fidelity monitoring was operationalized in three levels:

1. Deviation from intended intervention.
2. Deviation from components of the intended intervention.
3. Deviation from steps within components of the intended intervention.

Using this fidelity can be reported as: the intervention, component or step is not delivered at all; an intervention, component or step from another trial group is delivered instead; or an entirely different intervention, component or step is delivered. Reasons for the deviations are important and should be documented throughout the RCTs.

Outside of surgery, the Osteoarthritis Research Society International (OARSI) published a series

of papers beginning in 2015, focused on improving the conduct of clinical trials and implementation of evidence-based practices in osteoarthritis of the hip and knee. One paper, "OARSI Clinical Trials Recommendations: Design and conduct of clinical trials of surgical interventions for osteoarthritis" highlighted methodological challenges in the design and conduct of randomized trials of surgical interventions in osteoarthritis and proposed strategies to address them.[27] The paper focused on three broad areas: enrolment; intervention; and assessment. In the intervention category, surgical intervention design and implementation pose challenges relating to obsolescence, fidelity of intervention delivery, and adherence and crossover. Guided by Nelson's 5 step Intervention Fidelity Model[139], Intervention Fidelity was defined as "the extent that the intervention is delivered in an identical fashion to each subject.[27]"

The OARSI recommendations highlight the challenges of maintaining and monitoring fidelity in trials of surgical procedures.[27] In particular, the slight differences in individual participants' anatomy necessitate that different dissection planes, extent of lavage or control of blood loss, soft tissue modifications and other modifications be made by the surgeon in real-time. Echoing the recommendations of the IDEAL[3] framework and Blencowe et al.[169], OARSI recognizes that surgeons' overall experience and learning curves associated with the surgical intervention being investigated also present challenges for intervention delivery or fidelity assessment. To address these issues, OARSI recommends that surgical investigators should develop consensus on the precise surgical protocol, including the decisions that may be made due to particular intra-operative findings before the trial. These should be summarized in an intervention manual

or trial manual and be accessible to interventionists. Intra-operative filming may also facilitate fidelity assessment and identification of departures from the study or intervention protocol. In trials of novel procedures and materials, such as implantable devices or new instruments, surgeons with minimal experience with the procedure or equipment should not be involved in the study.[27]

## Intervention fidelity in rehabilitation

Another paper in the series by OARSI, "OARSI Clinical Trials Recommendations: Design and conduct of clinical trials of rehabilitation interventions for osteoarthritis," was published in 2015[170] targeting RCTs of the hip and knee, with applications to rehabilitation trials for other regions as well. The recommendations included considerations for entering participants into trials, conducting of trials, including enhancing and monitoring intervention fidelity and participant adherence, selection of outcome measures, and recommendations for statistical analyses and reporting of results. Thorough description of interventions, sufficient for replication, including intervention dosage (intensity, frequency, duration), procedures for tailoring interventions to individual participants, strategies to enhance participant adherence, training of treatment providers, and treatment fidelity methods was also recommended.[170] Intervention fidelity was not defined, nor were any fidelity modes or frameworks described or cited in the paper.

The importance of maximizing and monitoring participants' adherence to interventions in the trial was also emphasised, and a variety of methods to enhance adherence and minimize

dropouts, such as self-reported measures of exercise adherence, activity logs, accelerometers, encouragement phone calls from researchers or interventionists, and others were suggested. The OARSI task force recommendations also recognized that a high level of interventionist skill and expertise in applying and monitoring interventions maximizes treatment fidelity.[170] Training in the intervention should be provided to interventionists, along with supportive materials including study manuals, DVDs and in person or on-line training programs. It may also be necessary to verify that the interventionists have sufficient skill with the intervention following training before treating participants as part of the trial.

Although no fidelity models were referenced or described in the paper, the OARSI taskforce and recommendations sought to fill a gap in guidance for conducting and reporting of interventions, including intervention fidelity, in RCTs of rehabilitation in persons with osteoarthritis. Intervention fidelity had increasingly been addressed in behavioural and public health, education, nursing, and other disciplines over the previous decades, but far fewer studies guiding or monitoring intervention fidelity in physiotherapy and other rehabilitation disciplines have been published.[5,7,60] Intervention fidelity monitoring in clinical trials of rehabilitation interventions is infrequently described in published studies.[36,46,52,60,105,119]

Attention to intervention fidelity in physiotherapy research has been both recent (mostly within the decade), and limited.[60,118,171,172] For example, systematic reviews of physiotherapy in paediatric rehabilitation[173] and adult neurological rehabilitation[174] have found a lack of fidelity monitoring or measurement in most of the reviewed studies, limiting the conclusions that can be

drawn about intervention effectiveness from the included studies.[173] Morgan et al.'s 2016 systematic review of 36 trials of physiotherapy interventions for infants and children with cerebral palsy found no studies reporting either monitoring or measuring intervention fidelity.[173] Vaughan's 2015 scoping review of interventional and methodological paper for Neurodevelopmental Treatment, a frequently use treatment in physiotherapy for children and adults with neurological disorders, found that only one out of thirty-three eligible papers, including 12 experimental studies, reported monitoring interventionists' fidelity to the study manual and intervention procedures (achieving "adequate fidelity", predetermined as > 60% fidelity to study instructions).[175] Fidelity was defined as the degree to which a therapist implements an intervention under research conditions, with determination of causality depending upon establishing treatment fidelity. Lack of intervention fidelity monitoring, poor monitoring of participant adherence and treatment dosages, and inadequate reporting of intervention procedures sufficient for replication were identified as major limitations in stroke rehabilitation research overall.[174,176]

Other reviews of neurological rehabilitation (i.e., for stroke), including physiotherapy, occupational therapy, and speech-language therapy have found that intervention fidelity is rarely reported. An analysis of 182 stroke rehabilitation trials published during 2015 by Walker et al. (2017)[177], found that fidelity was reported for fewer than 10% of included trials. Of those reporting fidelity, most studies addressed a single aspect of intervention fidelity such as participant or therapist adherence (dose). Intervention fidelity was operationalised as whether interventions were delivered as intended, whether the interventions were modified to suit the

local context (setting or therapist) or individually tailored to the person being treated, and the methods used to maintain intervention fidelity during the trial. Items 8-12 from the TIDieR[150] checklist were used to assess publications for intervention fidelity.

More recent systematic reviews of interventions in speech-language therapy have also found that many studies fail to adequately plan, investigate, or fully report intervention fidelity.[178–180] For example, Brogan (2019) reviewed RCTs of speech-therapy interventions for post-stroke aphasia published between 2012 and 2017.[181] Defining intervention fidelity as "the degree to which the administration of a treatment corresponds to the specified protocol for the implementation of that treatment," Brogan assessed intervention fidelity across 42 eligible papers using the TIDieR[150] checklist and following the NIH-BCC[133] fidelity model. Only 9 of the 42 papers (21%) reported treatment fidelity monitoring or treatment fidelity processes. Only one paper (2%) contained every element of the recommended treatment fidelity areas. Overall, 37 (88%) papers addressed fidelity to study design, with the most reported element of intervention fidelity being the planned therapy dosage. This was most frequently reported as the number of minutes of therapy received. Enactment, or ensuring participants used intervention skills in daily life was the least addressed component of fidelity, with 2 (2%) papers including this. Additionally, little detail was provided for the make-up specific interventions, or their differentiation from comparison conditions.[181]

In other areas of physiotherapy and rehabilitation, such as cardiopulmonary rehabilitation, sports rehabilitation, orthopaedic or musculoskeletal physiotherapy and manual therapy,

intervention fidelity has generally been poorly addressed or reported.[52,182–189] A 2021 systematic review and meta-analysis of the efficacy and treatment fidelity of conservative treatment interventions for persons with shoulder pain assessed intervention fidelity in eligible papers using a modified version of Borelli's intervention fidelity checklist (NIH-BCC[39])[62]. Only 2 of the 10 included studies met the criteria for adequate intervention fidelity, and no study met all of the checklist's criteria for fidelity. The authors suggested that most studies included in their review could have been influenced by factors linked to intervention fidelity, including a lack of researcher training, lack of fidelity to protocols, and interventions not being performed as had been specified in the protocol or study manuals. The authors proposed that a lack of intervention fidelity could also help to explain why studies investigating similar interventions produced different results.[62]

A 2019 systematic review and meta-analysis of the efficacy and fidelity of clinical interventions used to reduce posterior shoulder tightness by the same authors, using the same methods, also found only 3 of the 13 included studies achieved good or better intervention fidelity.[190] The authors identified an overall lack of identifiable methods to ensure that the interventions were being performed as specified and delivered at the specified dose. Another systematic review and meta-analysis of physiotherapy (manual therapy) interventions for knee osteoarthritis found only one third of included studies had high levels of treatment fidelity.[52,191] The lack of fidelity monitoring or reporting in published physiotherapy trials is concerning, given the potential impact it may have on rehabilitation research and participant outcomes. Without monitoring and maintenance of intervention fidelity, there could be significant changes in the content and dose

of interventions or their active ingredients. The difference between a participant being treated for 30 minutes of physiotherapy, rather than the a study's set protocol of 20 minutes, could be substantial.[52]

Poor monitoring and reporting of intervention fidelity has been described in a variety of other papers in other rehabilitation disciplines. In occupational therapy, for example, Parham investigated *the* intervention fidelity of research studies in occupational therapy investigating sensory integration treatments, finding treatment fidelity was generally poor, compromising the conclusions that and be drawn about the intervention's effectiveness.[42] A lack of attention to intervention fidelity and poor consistency in the conceptualisation and measurement of fidelity has also been reported in trials of physical activity and exercise-based interventions.[45,46,192] A scoping review (2017) of 485 physical activity or exercise papers found that only 24 (5%) of published articles described intervention fidelity.[45] Analyses of exercise trials in meta-analyses of hypertension treatments have also found poor reporting of intervention fidelity and participant adherence across trials.[192]

A systematic review of intervention fidelity in physiotherapy by Toomey et al., followed the NIH-BCC Comprehensive Fidelity framework and used the updated 40 item fidelity checklist[138] to assess reported treatment fidelity among 22 studies of physiotherapist–led interventions to promote self-management for people with osteoarthritis and chronic low back pain.[171,172] Intervention fidelity was poor across the 22 included studies. The average fidelity score was just 37% out of a maximum score of 100% (all elements of the fidelity checklist

addressed).[172] Levels of fidelity varied within studies but were generally poor. The average level of fidelity across all studies of 35%, ranging from 12% to 63% , and no study achieving "high treatment fidelity," defined by the NIH-BCC as 80% or more on their checklist.[39] The "content of intervention" was reported adequately in all studies (100% fidelity score for that category) across all 22 studies. Considering the components of intervention fidelity identified in the NIH-BCC model, fidelity of treatment delivery was found to be 20% and fidelity of treatment receipt was 33%. Study design (53%) and treatment enactment (43%) were the most reported and highest scoring components .Training of providers was the category achieving the lowest fidelity score across studies (10%).

## Poor conduct, poor definition, or poor reporting?

The lack of focus on intervention fidelity within rehabilitation trials may be influenced by the lack of consensus for defining or conceptualising intervention fidelity or its components.[171,193] As well as the myriad of fidelity constructs and models that have been developed for psychology, behavioural health, education and public health, a number of fidelity constructs have been reported in rehabilitation papers. For example, Nelson (2004)[51], Hildebrand (2012)[194] , Persch (2013)[23] Breckenridge (2015)[195] Parham [42] Hand (2018)[143], and Mihee (2021)[60], among many others, have described or proposed variations on fidelity models and constructs adapted from psychology or behavioural health for rehabilitation research. The number of fidelity models, constructs, and definition in use in complex interventions in the physical domain has not been quantified, to the best of our knowledge. What is certain, is that there is large variation and poor consensus for what intervention fidelity should include, or how precisely to define it.

The poor reporting of intervention fidelity in physical complex interventions may also just reflect poor reporting, rather than poor conduct.[43,67,172,181] The conduct of a clinical trial and its reporting are two distinct concepts.[196] While intervention fidelity in physical complex interventions may not necessarily be poor, reporting of fidelity is. The 2010 updates to the CONSORT[148] guidelines, SPIRIT[26], TIDieR[150], and CERT[151] were developed to improve intervention reporting, including intervention fidelity. Nevertheless, systematic reviews of clinical trials in physiotherapy and other complex interventions find poor adoption of these standards, despite wide-spread dissemination over the past decade.[197]

Systematic reviews of the quality of intervention reporting in rehabilitation have found that many clinical trial papers do not report intervention details sufficiently to allow for reproducibility, or fail to completely or acutely comply with the reporting recommended by the SPIRIT[26], TIDieR[150], and CERT[151] checklsits.[198–202] For example, McCambridge et al. (2021) evaluated 140 RCTs published in six leading physiotherapy journals between 2000 and 2018.[202] The authors assessed each paper with the TIDieR[150] checklist and the Physiotherapy Evidence Database (PEDro)[203] quality scale. They identified control and treatment interventions, exercise-based interventions, and the area of physiotherapy practice. A 20% difference in the TIDieR[150] or PEDRO[203] scores between studies published in 2000 and 2018 was considered a meaningful change. One hundred forty papers that met selection criteria, including 225 interventions. Their analyses found that reporting of interventions in physiotherapy RCTs had not meaningfully improved since the implementation of the TIDieR[150] guidelines, with approximately 42% of all items checklist not fulfilled in the 2018 papers. Control interventions were reported less

completely than the experimental interventions, although reporting for those did not change meaningfully between 2000 and 2018. Only the number of treatment sessions was reported more completely in 2018 than in 2000. Other systematic reviews in physiotherapy, speech-language therapy, and cardiopulmonary rehabilitation, for example, also report a lack of detail necessary to allow accurate evaluation and replication in trial publications.[181,198–202,204]

### Why so poor? Possible barriers to intervention fidelity and fidelity reporting

The reasons for under-reporting of fidelity monitoring in trials of complex intervention are various and complex. Poor reporting of intervention fidelity may be influenced by editorial issues such as word count limitations in scholarly journals and a lack of journal editorial requirements to report intervention fidelity.[1,12,56] While the CONSORT[148] guidelines, and the SPIRIT[26], TIDieR[150], and CERT[151] checklists include items for intervention fidelity, they do not offer specific guidance about how to actually report intervention fidelity in clinical trial papers.[1,172] Surveys of researchers, trialists, and interventionists in complex interventions have found that the lack of practical guidance, along with lack of criteria specifying acceptable levels of intervention fidelity, difficulty quantifying fidelity data, and a lack of perceived importance are barriers to addressing or reporting intervention fidelity.[1]

Several other potential barriers to treatment fidelity and fidelity monitoring in complex intervention clinical trials have been proposed, including the potential increase in time, resources and cost required to monitor, measure, and report fidelity.[1,11,36,45,48,172,205] Lack of knowledge and need for further training and education around intervention fidelity are also

perceived barriers to enhancing or monitoring intervention fidelity in clinical trials.[1,152] Other reported barriers include potentially increased participant and provider fatigue caused by fidelity or adherence monitoring record keeping or frequent intervention protocol reminders.[52,206]

Additionally, difficulty standardising complex interventions, with their multiple interacting components, and their delivery have been identified as a barrier to intervention fidelity in their clinical trials.[207] Pragmatic randomised clinical trials, in particular, offer practitioners "considerable leeway in deciding how to formulate and apply[208]" interventions, making strict standardization of interventions more difficult and inferences about the relationships between intervention components and treatment effects more complex. [5,7,13] Pragmatic trials are also designed to evaluate the relative effectiveness of interventions under real-life[209] conditions, with diverse clinical populations,[210] and against usual care interventions, further challenging how fidelity is monitored or measured.[211,212]

Interventionists in rehabilitation (for example, in physiotherapy or occupational therapy) may also perceive that treatments in their discipline are highly individualized by nature, and not well suited for detailed operationalisation or fidelity measures.[143] Interventionists may also feel that fidelity monitoring sets stringent criteria for intervention delivery that limit flexibility and their ability to adapt or tailor interventions to individual participants based on the interventionists' clinical experience.[1,213] A strict focus on intervention fidelity may, therefore, restrict interventionists "therapeutic freedom," restricting interventionists' professional autonomy and imperilling intervention outcomes.[1,52,172,214,215] Increasingly, however, careful, pre-planned

adaptation of interventions to individual participants and settings is seen as an important part of intervention fidelity, enhancing intervention effectiveness and the translation of interventions into evidence-based practice, rather than as opposing concepts.[152,216,217] Interventions can be designed to allow for flexibility within fidelity, allowing tailoring to individual patients while still remaining faithful to their treatment theory or the study's program theory and avoiding proscribed behaviours, particularly in pragmatic trials.[152,216–218]

Understanding of the possible reasons for poor intervention fidelity during clinical trials, the experiences and attitudes of stakeholders involved in these trials, and their contribution to treatment fidelity remains insufficient, however. Nevertheless, there is also growing recognition that the benefits of ensuring fidelity and addressing knowledge gaps in intervention fidelity in complex interventions outweigh the potential costs and investments in time and labour.[48]

## Evidence gaps explored in this thesis

Throughout the literature review that informed this chapter (Chapter 1), several knowledge gaps in the evidence for intervention fidelity were identified. Four primary evidence gaps will be addressed in this thesis:

1. Fidelity constructs: A number of fidelity frameworks have been developed and published for use in behavioural health, public health and education, but evidence-based guidance for planning, enhancing, or monitoring intervention fidelity monitoring in clinical trials of complex interventions in the physical domain (surgery and rehabilitation, broadly defined) is

rare.[40,45,63,219] A comprehensive fidelity framework for use in the physical domain could support the design and conduct of more robust clinical trials and support the replicability of evidence-based interventions in physical complex interventions.[16,117,220–222]

2. Defining fidelity: Very little guidance was found for assessing whether intervention fidelity was supported, maintained, monitored, or measured in clinical trial *publications*. Definitional inconsistencies and the number of varying conceptual interpretations of intervention fidelity undermine identification of its core components and may foster inconsistent application of methods to enhance or monitor intervention fidelity in complex intervention clinical trials.[1,12,40] The large number of terms and definitions used for intervention fidelity may also make it difficult to identify fidelity monitoring in clinical trial papers. Those who develop, implement and study complex interventions have no common language by which they can make assessments and develop robust methods to support intervention fidelity. [10,40,63,117]

3. Fidelity and treatment effect estimates: It has been proposed that poor intervention fidelity in clinical trials can dilute or diminish treatment effect sizes due to contamination bias and extraneous variables, and trials with low fidelity would need significantly larger sample sizes to compensate for this.[53] However, very few papers have provided evidence that improved intervention fidelity leads to improved statistical power or can influence the magnitude or direction of treatment effect sizes in clinical trials. The effect of poor intervention fidelity on clinical trial treatment effects remains theorized, rather than quantified.

Research is needed to evaluate whether, and how much, intervention fidelity effects treatment

effect estimates in clinical trials of complex interventions.[52] While a range of factors biasing treatment effects, such as randomisation and blinding, have been examined in meta-epidemiological research, no meta-epidemiological studies have investigated intervention fidelity in clinical trials, to the best of our knowledge. The effects of biases in RCTS can be unpredictable, masking treatment effects or reversing their direction, shifting treatment effects toward or away from the null,[73] particularly when the results of RCTs are pooled in meta- analyses.[223] Determining the size and direction of intervention fidelity's influence on treatment effect estimates in complex intervention clinical trials can improve interpretation of their results and their appropriateness to change practice. In this thesis, the term treatment effect will refer to the estimated causal effect that the experimental intervention (independent variable) has on the primary outcome (dependent variable) of the study.

4. Pragmatic trials: Much of the research identified in this review explores intervention fidelity in the context of explanatory (efficacy) trials. Their tightly controlled nature and focus on examining interventions under the best possible conditions to show an effect on outcomes careful monitoring and documentation of intervention fidelity. However, pragmatic randomised clinical trials are designed to evaluate the relative effectiveness of interventions under real-life[209] conditions, with diverse clinical populations,[210] and against usual care interventions.[211,212] Key aspects of intervention delivery are likely to be less tightly controlled in pragmatic trials, creating challenges for assessing and maintaining intervention fidelity.[210,224,225] Pragmatic trials can also be challenging for intervention fidelity, as they often monitor, rather than ensure, fidelity.[226] The applicability of fidelity constructs in pragmatic trials has not been explored extensively.

Fidelity may look different in a pragmatic trial compared to an explanatory trial, for example, and may depend on the goals of the trial as well as the kind of interventions being investigated. Detailed explorations of intervention fidelity concepts, or their applicability, in pragmatic trials of physical complex interventions are yet to be published.[216,224]

## Aims of this thesis:

1. Quantify the number of fidelity constructs, models, frameworks, or monitoring reported in empirical and methodological papers reporting the delivery of complex interventions in the physical domain: surgery, rehabilitation (physiotherapy, occupational therapy, speech and language therapy), and exercise or physical activity interventions.
2. a. Synthesize a conceptual framework for intervention fidelity in clinical trials of complex interventions in the physical domain.

   b. Guided by the conceptual framework, develop a fidelity checklist to allow transparent, reproducible assessment of intervention fidelity in RCTs of physical complex interventions, facilitating aim 3.
3. a. Determine whether intervention fidelity influences the magnitude of treatment effect estimates derived from RCTs of physical complex interventions.

   b. Determine in which direction treatment effect estimates are biased by poor intervention fidelity.

   ### Hypotheses:

   a. Poor intervention fidelity quantifiably biases treatment effect size and precision in RCTs of physical complex interventions.

b. The direction of the bias is away from the null. High degrees of intervention fidelity are associated with a larger treatment effect estimates and greater precision (smaller standard errors and more narrow confidence intervals[70,227]) while poor fidelity supresses treatment effects and reduces precision in complex intervention RCTs.

4. Apply the CONSIDER framework to explore intervention fidelity and participant adherence in the ACL Surgery Necessity in Non Acute Patients (ACL SNNAP) study to:

a. Determine if CONSIDER be used to evaluate the fidelity of a complex intervention pragmatic trial.

b. Describe the types of fidelity and adherence encountered in an ongoing clinical trial

c. Describe how fidelity and participant adherence can affect the analysis and interpretation of ACL-SNNAP's results

d. Glean insights about the results of the meta-epidemiological study through the experience of trial conduct and reporting in the ACL-SNNAP trial.

## Key terms:

In this thesis, an integrated definition of **intervention fidelity** is used (chapter II), in which intervention fidelity is an umbrella concept encompassing two distinct but related and interacting components: intervention fidelity and participant adherence. Both are essential for a clinical trial to be faithful to its protocol, and both can influence study outcomes, individually or together:[3,19–21]

1. Participant Adherence (adherence) focuses mainly on the actions of the participant. Adherence encompasses both whether participants accept and initiate the intervention allocated, and how well they comply with the prescribed, allocated intervention. For example, this could be represented by measuring whether participants attended the required number of therapy sessions, the frequency of intervention sessions or frequency participants performed intervention activities (for example, completed a home exercise a certain number of times per week).
2. Bias: Bias will be defined as per the Cochrane Handbook, version 6.2: "A systematic error, or deviation from the truth, in results or inferences."[70] The magnitude of biases can vary and they can lead to either "underestimation or overestimation of the true intervention effect."[70]
3. Intervention delivery fidelity (fidelity) focuses mainly on the actions of the interventionist. It encompasses the quality of intervention delivery or performance within the trial, and reflects the correspondence of interventions delivered in the trial with the intervention specified in the study protocol, or in accordance with study procedures, treatment manuals, etc.
4. Treatment effect estimate: In this thesis, the term treatment effect will refer to the estimated causal effect that the experimental intervention (independent variable) has on the primary outcome (dependent variable) of the study.

Chapters II, III, IV in this thesis focus mainly on fidelity and actions of the interventionists, while chapter V focuses mainly on adherence and actions of the participants.

## General methods overview

The following subsection lists the primary aims and methods for each chapter.

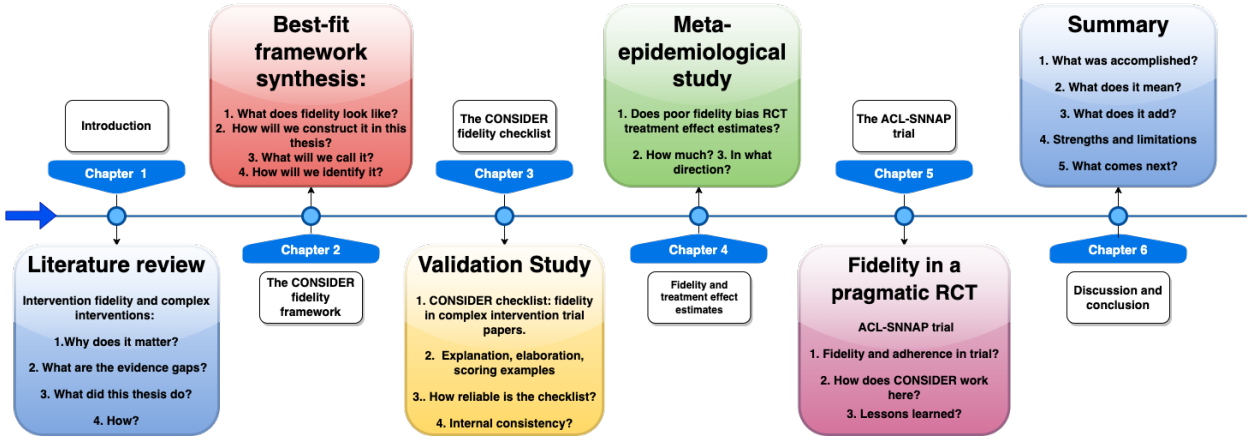



**Figure 1.3:** Thesis flow diagram

### Chapter I:

**Main aim:** Describe the development of intervention fidelity in complex intervention research, identifying the current state of knowledge and evidence gaps, and the context of this thesis.

**Method**: Literature review

### Chapter II:

**Main aims:** To quantify and describe fidelity models, constructs, and definitions reported in complex intervention literature and construct a comprehensive intervention fidelity model from which to investigate intervention fidelity in this thesis. A second aim was to construct a transparent, reliable fidelity checklist to allow transparent, reproducible assessment of

intervention fidelity in RCTs of physical complex interventions for the meta-epidemiological study in chapter 4.

Methods: Systematic review, Best-Fit Framework Synthesis, Thematic analysis.

Chapter III:

Main aims: To assess the internal consistency and inter-rater reliability of the CONSIDER checklist derived in Chapter 2 and gather participants' experiences using the checklist to identify items needing further development, based on user feedback.

Methods: Explanation and elaboration paper, and a reliability and pilot study with 16 participants each assessing three randomly selected papers with the CONSIDER checklist.

Chapter IV:

Main aims: To determine whether intervention fidelity influences the magnitude of treatment effect estimates derived from RCTs of physical complex interventions and determine in which direction treatment effect sizes are biased by poor intervention fidelity.

Methods: A systematic review and meta-epidemiological study intervention fidelity in meta-analyses of controlled trials physiotherapy, occupational therapy, speech-language therapy, exercise, and physical activity. By contrasting the results of trials with a characteristic of interest with the results of trials with without that characteristic, the meta-epidemiological study provided estimates of the average bias associated with poor intervention fidelity.[79,80]

Chapter V:

Main aims: To apply the CONSIDER framework in an ongoing pragmatic trial, the ACL Surgery Necessity in Non Acute Patients (ACL SNNAP) study. This chapter investigates the framework's

feasibility and applicability in a contemporary pragmatic trial, describes the types of fidelity and adherence the trial, including how fidelity violations influence interpretations of the trial's results, and gleans insights about the results of the meta-epidemiological study through the experience of trial conduct and reporting in the ACL-SNNAP trial

Methods: Fidelity assessment, quantitative analyses, thematic analysis, and synthesis.

Chapter VI:

Main aims: To discuss the results and implications of the thesis, along with its strengths, limitations, and future directions.

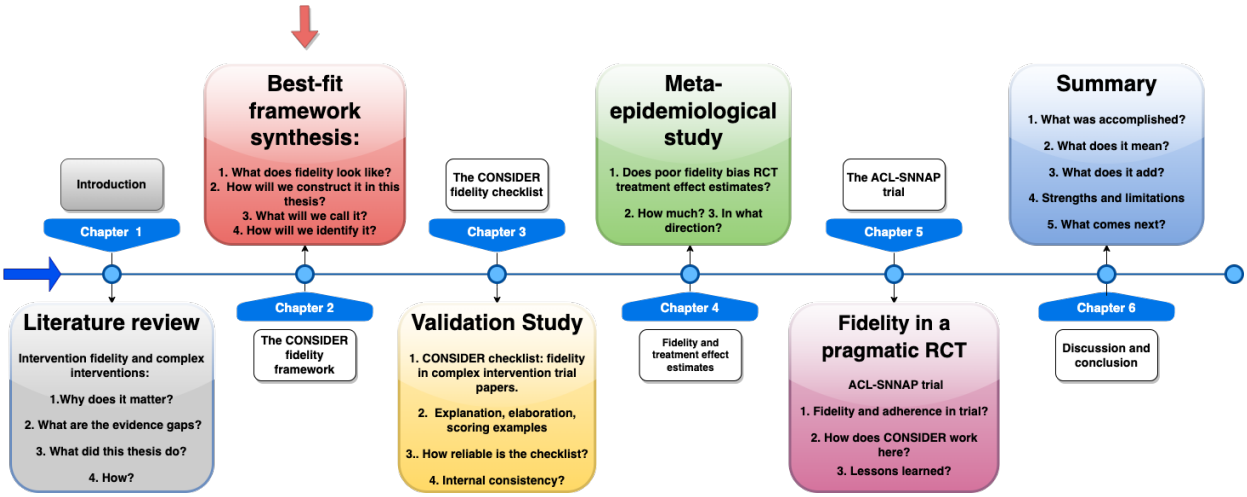
Appendices: Include search strategies and results, checklists, and screening results, and supporting materials for chapters 2-6.

Chapter II: The CONSIDER Intervention Fidelity Framework for Complex Interventions: A "Best-Fit" Framework Synthesis

## Chapter II:

## Chapter Summary

In this chapter, a systematic review followed the 'best-fit' framework synthesis (BFFS) approach to develop an empirically based, conceptual framework to guide investigation of intervention fidelity and define key terms in this thesis. The resulting CONSIDER (**CO**mplex i**N**terventions trial**S** f**IDE**lity f**R**amework) framework conceptualizes fidelity in terms of intervention and trial design, delivery, and receipt, and recommends processes to support and monitor intervention fidelity in clinical trials. An integrated definition for intervention fidelity in this thesis was also synthesised and presented in this chapter.



## BACKGROUND

A variety of frameworks have been developed to understand and monitor intervention fidelity in clinical trials, but these vary in content and focus. There is also little consensus on how best to define intervention fidelity or categorize its key components, further complicating efforts to enhance or monitor intervention fdielity.[10,11,40,105,228] Those who develop, implement and study complex interventions have no common language by which they can make assessments and develop robust methods to support intervention fidelity.[10,40,63,117]

Updates to the CONSORT[148] reporting guidelines (2010), The Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT)[26] statement (2013), Template for Intervention Description and Replication (TIDieR) Checklist[150] (2014), and the Consensus on Exercise Reporting Template (CERT)[151] (2016) were developed to improve intervention reporting and include items related to fidelity. However, they provide limited guidance for reporting strategies to maintain participants adherence to their assigned intervention or document modifications to intervention allocation, and no guidance on strategies to monitor or support the quality of intervention delivery. These checklists also only provide guidance on what should be reported and were not intended to offer researchers a practical guide on how to approach intervention fidelity in study design or conduct.[146] What was *not* done in a trial cannot be reported.

Treatment effects observed in trials will not be reproducible unless well-described procedures facilitate and maintain intervention fidelity in clinical trials.[16,221,222] Published fidelity frameworks

have been developed for use in behavioural medicine, public health and education, but evidence-based guidance for fidelity monitoring in clinical trials of physical complex interventions such as surgery, physiotherapy, and rehabilitation is rare.[40,45,63,219] A comprehensive fidelity framework for clinical trials in domains involving physical complex interventions, such as physiotherapy and rehabilitation (physical medicine and rehabilitation, broadly defined) was needed as a framework from which to investigate intervention fidelity in this thesis.

A conceptual intervention fidelity framework is made up of a set of guidelines or recommendations detailing a combination of strategies and methods to assess, enhance and evaluate intervention fidelity at different stages of an intervention's implementation during a clinical trial.[229] In this chapter, the development of the **CO**mplex i**N**terventions trial**S** f**IDE**lity f**R**amework (CONSIDER) framework, a conceptual framework for intervention fidelity in clinical trials of complex interventions in the physical domain, is described. Further development of this framework through broader input from a wider range of stakeholders could support the design and conduct of more robust complex intervention clinical trials. [1,117,220]

## Objectives

1. To provide a description of fidelity constructs, models, frameworks reporting the delivery of complex interventions in controlled trials in physical domains of health care: physiotherapy, surgery, physical medicine, rehabilitation, occupational therapy, speech and language therapy, exercise, and physical activity promotion.

2. To conduct a best-fit framework synthesis of fidelity constructs to inform a conceptual fidelity framework from which to investigate fidelity in trials of complex interventions in physical domains in this thesis.
3. To produce an integrated, empirically based definition of intervention fidelity for use in this thesis.

## METHODS

### Reporting standards

The systematic review and framework analysis are reported along the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [230] and **EN**hancing **T**ransparency in **RE**porting the synthesis of **Q**ualitative research (ENTREQ) statement guidelines.[231]

### Registration

The protocol was registered on PROSPERO,[232] (https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42019135957) following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Protocols (PRISMA-P) guidelines.[230]

### Systematic Review Design

The review followed a Best-fit Framework Synthesis (BFFS) approach to develop a comprehensive fidelity framework based on existing evidence (Fig. 2.1).[233] Best-fit framework syntheses use initially-identified conceptual models found in published literature to serve as a

platform from which a new framework can be developed through thematic synthesis of evidence from empirical studies.[233–236] The approach is augmentative and deductive, building iteratively on the existing models.[236] We chose to create a new intervention fidelity framework though BFFS as, to the best of our knowledge, no intervention fidelity frameworks had previously been developed to guide fidelity in clinical trials of complex interventions in the physical domain.

Very detailed academic methodology was needed to construct the framework (part one). To enable it to be useful, it also had to be distilled into useful portions, headings, and silos that can be easily illustrated and recalled. The review is therefore divided into two stages (figure 2.1).

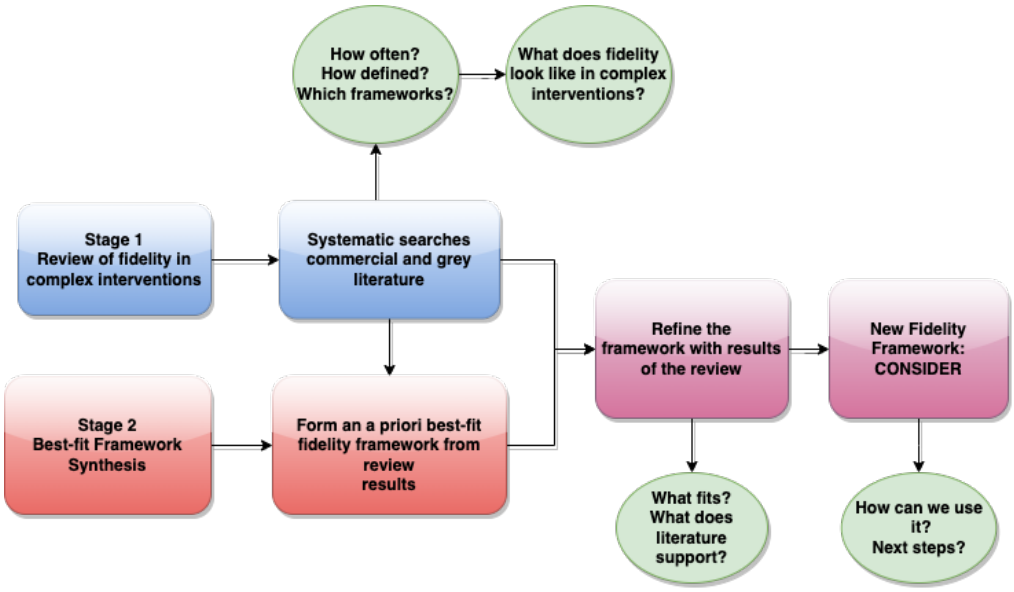



**Figure 2.1:** Stages of the review and best-fit framework synthesis

### Stage 1: Create the platform ("Best-fit" framework) on which to build the new framework

Once the research question or aim was determined, the creation of the *a priori*, best-fit

framework (a platform from which to build a new framework) was conducted in parallel with systematic searches for papers to be included in the review and synthesis . These two "strands" then joined at the framework synthesis stage (stage 2).

### Stage 2: Build on the platform with empirical evidence to make a new framework

In stage 2, fidelity frameworks, theories, models and concepts identified in stage 1 were thematically synthesized and reciprocally translated through BFFS (figure 2.2) [233,236] to develop an integrated, conceptual fidelity framework for complex interventions in the physical domain. This new framework can subsequently be honed and validated in future empirical research.

### Stage I: Systematic review

### Search Strategy

As we were searching for theoretical frameworks, the search strategy was informed by the BeHEMoth (Behaviour of interest, health context, exclusions, and models or theories) method.[234,237 235]

Be: Treatment or intervention fidelity, adherence, integrity, compliance, concordance, implementation, and related concepts.

H: Complex interventions in the physical domain in healthcare: physiotherapy, occupational therapy, speech-language therapy, exercise or physical activity interventions, surgery.

E: Statistical or economic models, models of care. Non-healthcare interventions.

MoTh: Model, theory, theories, framework, concept, conceptual, construct or strategy.

Keywords and MeSH terms related to fidelity (table 2.1) were used in Pubmed, Embase, CINHAL, Scopus and Google Scholar (appendix II figure 2.1). PROSPERO was searched for ongoing or recently completed systematic reviews. Searches were limited to papers published from 2005, a year after the publication of the NIH Behaviour Change Consortium (NIH BCC) recommendations for conceptualizing and enhancing treatment fidelity.[133] No language limits were imposed on searches. To optimize coverage and ensure literature saturation, grey literature searching was conducted on Scopus, Google Scholar, and PROSPERO. Grey literature includes a wide range of data sources not typically captured in searches for commercially published literature, including clinical trial and protocol registers such as clinicaltrials.gov[238], conference proceedings, government reports, not-for-profit or non-governmental agency reports, international bodies' reports or position papers (e.g., the World Health Organisation).[239] The reference lists of included papers were also searched for additional papers. Study protocols or trial registrations were also searched for and reviewed when they were available.

| Fidelity: Key words and related terms |
| --- |
| Treatment, intervention, procedure, trial, according, adherence, agree, agreement, conceptual, concordance, consistent, compliance, delivery, delivered, directions, enactment, evaluation, fidelity, framework, implementation, Indicators, integrity, intended, instructions, manual, manualized, model, monitored, monitoring, process, protocol, similar, similarity, specified, standard, strategy |

**Table 2.1:** Fidelity keywords and related terms

## Inclusion criteria

The aim of the systematic review was to identify as many relevant studies as possible and reduce the risk of missing potentially eligible studies, maximizing sensitivity rather than precision.

Eligible papers were complex interventions empirical research, review, or theoretical papers including terms related to fidelity anywhere in the paper, either as a main focus or component (for example, as an analysis within a trial or process evaluation).

### Exclusion criteria

Papers not describing fidelity, or a related term, were excluded (table 2.1). Papers investigating fidelity outside of complex interventions in the physical domain in healthcare were ineligible. These included papers for public-health interventions, such as smoking cessation or reproductive health interventions, interventions in education, such as reading proficiency, or in psychology, such as interventions for compulsive behaviour or depression.

### Study selection

Potentially relevant citations, their abstracts and full-texts were screened against the review's inclusion-exclusion criteria independently by the first author (AP) and a second author (DH). Disagreements were resolved by consensus. As the BFFS aimed for conceptual saturation and generalizability, rather than statistical generalizability, we aimed for study selection that was purposive rather than exhaustive.[240,241] A criterion-sampling approach to purposive sampling was used to select papers from the systematic review for framework synthesis.[241,242]

### Data Collection and Extraction Process

Citations, abstracts, and full text articles were managed digitally with Mendeley. A tailored, electronic data extraction form was created and calibrated for this review. We extracted papers' bibliographic information, field, professional discipline or context, study design or paper format

(e.g., methodological paper), descriptions or definitions of fidelity, fidelity monitoring or support methods, and fidelity model or framework if reported (Appendix II table 2.1).

### Quality assessment

Little consensus exists for the feasibility and utility of quality assessment in framework analysis or qualitative synthesis.[243] As this systematic review sought to synthesize fidelity frameworks and models, rather than the outcomes or effects of processes or procedures in empirical work, no quality assessment of papers was performed in this review.

### Data Synthesis

Objective 1: The results of the systematic review were described narratively and with simple descriptive statistics. No quantitative synthesis was undertaken.

Objective 2: Fidelity models or frameworks selected through purposive sampling were thematically synthesized with NVivo12 software.[233,236]

### Stage II. Build on the platform to make a new framework (Best-Fit Framework Synthesis)

With the results of the systematic review, the "Best-fit Framework Synthesis (BFFS) [233,236]" method was used following a series of predetermined steps (Figure 2.2).[244] As an overview, once an initial "platform" best-fit framework was constructed, fidelity data (fidelity descriptions, definitions, monitoring and enhancement) extracted from papers identified in the systematic review was thematically analysed and coded against the best-fit framework to expand or reduce it with the data to creating a new fidelity framework.[233–236]

Thematic analysis followed the guidance and recommended steps of Saldana's (2016) Coding Manual for Qualitative Researchers, and the Streamlined Codes-to Theory Model for qualitative inquiry.[245] First-cycle coding followed a descriptive (topic) coding method, generating a categorized inventory of codes identified and themed through careful reading and reflection on the data.[246] In second-cycle coding, pattern coding was used to reanalyse and reconfigure the first-cycle codes categorically, thematically, and conceptually into a smaller number of themes (pattern codes) to develop an understanding of the corpus of data and relationships between its components. [247]


**Figure 2.2:** Best-Fit Framework Analysis' Steps: Adapted from Booth, et al, 2014[244]

## Best-fit framework: steps

### Step 1: Identify the review question or aim.

Aim: to synthesize a conceptual framework for intervention fidelity in clinical trials of complex interventions in the physical domain (physiotherapy, surgery, rehabilitation).

### Step 2: Systematically identify relevant research and "best-fit" frameworks, models, or theories.

Comprehensive, systematic searches were conducted to identify as many relevant publications addressing the question as was feasible (Stage I). In a parallel process, these papers were reviewed to identify a best-fit "platform" fidelity models (Step 2b).[244]

Step 3: Generate the *a priori* framework from identified publications through thematic analysis. This *a priori* "best-fit" framework serves as a platform or starting point that will be built on with fidelity data from eligible papers to synthesise a new fidelity framework in the next steps.

### Step 4: Code evidence from included studies against the best-fit framework.

Once the BFF was constructed, passages describing fidelity in eligible papers were extracted as direct quotations and imported into in NVivio12. These quotations were systematically reviewed against the fidelity concepts categorised in the BFF, feeding them into the BFF theme they best represented.[248] This "thematic coding" was performed and critically reviewed by two reviewers (AP and DH) and any discrepancies or disagreements were resolved by consensus.

### Step 5: Create new themes.

When fidelity data from the extracted quotations did not translate well into any of the BFF concepts (themes) or were applicable to more than one, new themes were created through interpretive, inductive secondary thematic analysis[233,249] and reciprocal translation (Step 5).[250] The criterion for forming a theme was that at least two quotations addressing the same concept were identified, and they did not correspond to an existing fidelity theme. The BFF was expanded upon, reduced, or added to iteratively as successive studies were analysed and data was coded, ultimately creating a revised fidelity framework (Step 6).[233,236,251]

### Step 6: Produce a new framework composed of *a priori* and new themes.

Consensus was sought among the reviewers (authors) on which of the BFF fidelity themes were supported, and whether quotations extracted from eligible papers mapped onto a pre-existing theme, or themes mapped onto each other (could be collapsed) through reciprocal translation. This, and steps 4-5 resulted in a new framework, included some fidelity themes from the BFF and new themes derived from fidelity data extracted from the results of the systematic review.

### Step 7: Revisit evidence to explore the relationship between fidelity themes or concepts.

The resulting finalized list of fidelity themes was used to create a new, integrated fidelity framework for physical complex interventions. An interpretation of the content of the fidelity themes, and relationships between them, is presented in the results section.

## RESULTS

### Systematic Review

Searches produced 2857 records, 361 of which were screened for eligibility by full text (figure 2.3). Of these, 269 met inclusion criteria in the systematic review. One hundred forty-one papers defined or described fidelity, fidelity components or a fidelity framework, meeting criteria for informing the framework synthesis.

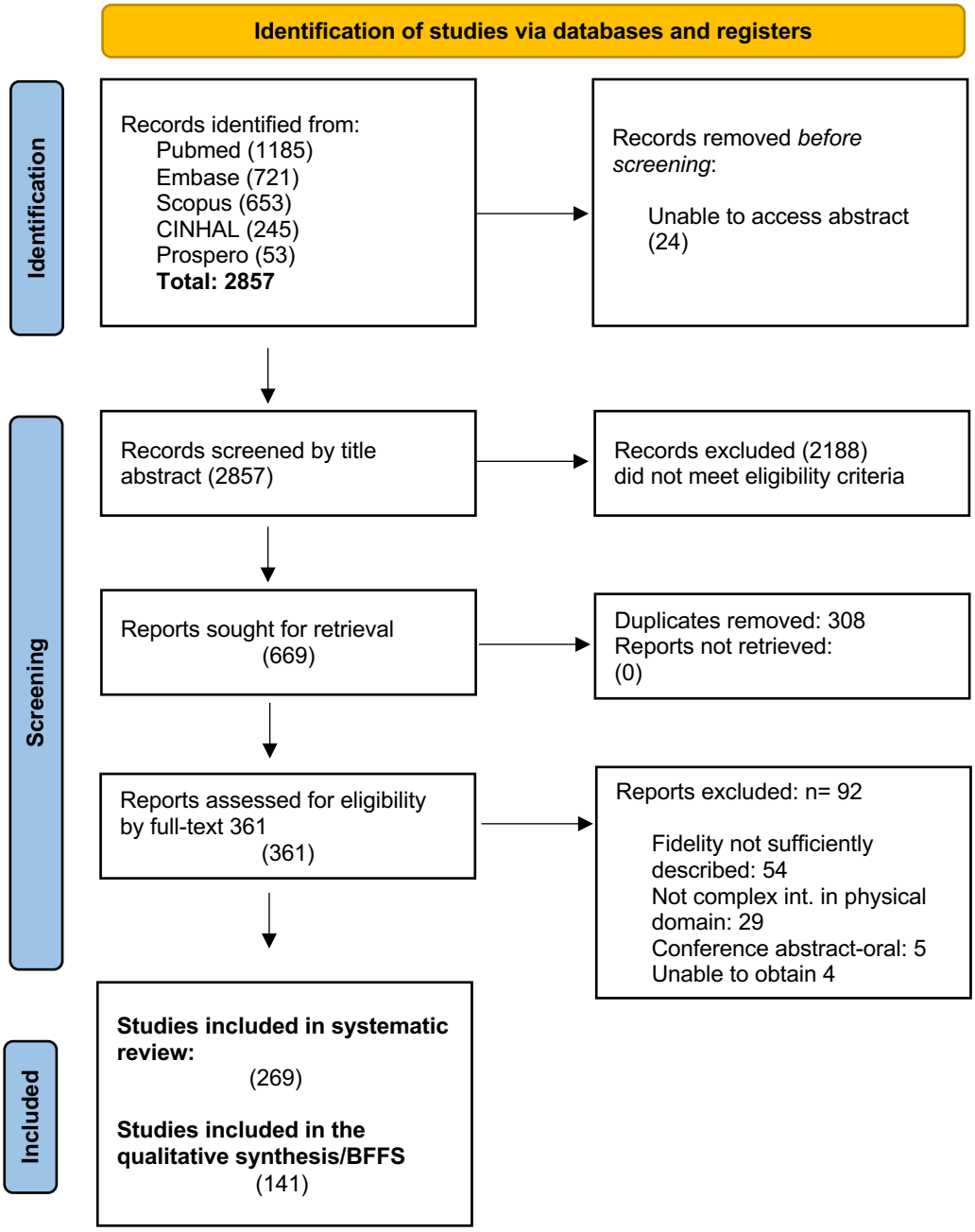



Figure 2.3: PRISMA flow chart

## OBJECTIVE 1: Fidelity in complex interventions

The greatest number of eligible papers (73) came from physiotherapy literature (table 2.2). The least represented specialty was surgery, with 16 papers, including three randomised controlled trials (RCTs) of a surgical procedure[252–254], two surgical trial protocols[255,256], one pilot [257], three trials of peri-operative surgical care or procedures[258–260], two systematic reviews of methodological reporting in surgical placebo-controlled trials[90,261], two guidelines on the conduct of surgical trials[16,27], two methodological papers[261,262], and one set of reporting standards.[169]

| Clinical discipline | Eligible papers | Study design or type of paper | Eligible papers |
|---|---|---|---|
| Physical Therapy | 73 | Protocol paper | 51 |
| Physical activity interventions | 66 | Randomized, parallel group trials | 47 |
| Rehabilitation (multidisciplinary) | 37 | Feasibility studies and trials | 35 |
| Occupational Therapy | 35 | Fidelity/Implementation studies, process evaluations | 33 |
| Speech Language Pathology | 25 | Literature or narrative reviews | 28 |
| Cardiopulmonary Rehabilitation | 17 | Methodological papers | 27 |
| Surgery and surgical care | 16 | Systematic reviews and Meta-Analyses | 23 |
| | | Case and pilot studies | 19 |
| | | Uncontrolled studies | 6 |
| Total | 269 | Total | 269 |

Table 2.2: Included studies according to discipline and study design/type

## OBJECTIVE 2: A best-fit framework synthesis

### BFFS Steps 1-3: Fidelity frameworks:

Steps 1-2b identified 28 fidelity frameworks or constructs across 108 papers. The two most frequently cited or applied fidelity frameworks were the National Institutes of Health Behaviour Change Consortium (NIH-BCC) Comprehensive Treatment Fidelity Framework[39] as described by

Bellg[133] or Borelli[39] (60 papers cited or applied), and the Comprehensive Intervention Fidelity Guide developed by Gearing[40] (23 papers cited or applied (table 2.3). The next most cited (19 papers) was the Implementation Fidelity Framework by Carroll et al (2007)[2]. The NIH-BCC and Gearing frameworks were thematically analysed to identify commonalities and differences between them and generate fidelity themes supported with examples extracted from the original paper (Table 2.3). These themes generated the best-fit framework (BFF) that served as a starting point, or platform, from which to build a new fidelity framework with fidelity data extracted from eligible papers (Step 4).[233,236,244,251]

The BFF was comprised of five fidelity themes (table 2.3).[63] Each theme was supported with a definition and supporting exemplars from the original papers. Fidelity data extracted from eligible papers was later coded against these themes in step 4.[233] The resulting BFF described fidelity on both a theoretical and an operational level (table 2.4).[63] The theoretical level addresses construct fidelity, or interventions' faithfulness to their underlying theoretical basis. Similar to Meehl and Cronbach's[263] concept of "construct validity," encompassing relations between hypothesized entities and processes and observed effects, the theoretical level encompasses interventions' design, proposed mechanisms of action, and their relation to the interventions' core effects.[40,264] The operational level encompasses interventions' fidelity in terms of the quality of intervention delivery, receipt (dosage) of interventions, participant engagement, and participant enactment, or whether participants apply intervention skills in their daily lives outside of treatment sessions.[40]

| NIH-BCC Comprehensive Fidelity Framework[39] | Comprehensive Fidelity Guide by Gearing et al, 2011[40] | Themes for Best-fit framework (sample sources) and exemplar text |
|---|---|---|
|  |  |  |
| **Study design**<br><br>Does study adequately test its hypotheses in relation to underlying theoretical and clinical processes? Are interventions' active ingredients fully operationalized? | **Intervention design**<br><br>Intervention's framework and elements essential to the design of a trial, its evaluation or replication. Includes a programme model, treatment manual. | **Intervention and study design**[2,12,268–271,40,51,67,153,181,265–267]<br><br>Treatment theory and mechanism of action, identification of active ingredients, differentiation from control or comparator interventions.<br><br>Identification of elements of the design of the trial investigating the intervention, including the trial's programme model and development of study procedures and/or treatment manuals.<br><br>**Exemplars**<br>1. "Conceptualization of a theoretical framework underpinnings treatment allows the key components that underlie the intervention to be measured and reported.[181] "<br>2. "Active ingredients: These active ingredients may be based largely on theory, or they may be based on empirical evidence. Active ingredients most typically include specific treatment targets, the therapeutic techniques, and the requirements for dosage (e.g., highly concentrated exposures several times per week). In combination, the active ingredients describe how and why the intervention brings about predicted outcomes.[266]"<br>3. "To achieve treatment fidelity, both the degree of integrity and differentiation are critical. Treatment differentiation requires that the experimental intervention and the intervention comprising the control condition differ in the intended manner. Treatment differentiation is determined by trial design and defined *a priori* by the investigators.[267]"<br>4. "Study design focuses on the methodological processes that ensure the study adequately assesses the proposed hypotheses in relation to a theoretical framework.[272]" |
| **Provider Training**<br><br>Is there standardizing training between providers, ensuring providers are trained to criterion, and is there monitoring and maintaining of provider skills over time? | **Intervention training**<br><br>Intervention fidelity requires adequate training and supervision of interventionists. Training elements include interventionist differences, such as levels of skills, education, experience, and implementation styles. | **Interventionist training**[2,12,67,153,172,213,271,273,274]<br><br>Interventionists' training and skill with delivering the intervention during the clinical trial, and their competence to deliver the intervention.<br><br>**Exemplars**<br>1. "Treatment fidelity related to the training of interventionists assures that those implementing the intervention were adequately prepared to do so and implemented all aspects of the treatment as intended.[273]"<br>2. "A specific intervention cannot be delivered until those delivering it have learned to do so in a standardized way.[275]"<br>3. "Thus, a team that has multiple protocol violations would be said to have low training fidelity, which would increase variability with which the protocol is administered.[23]" |
| **Treatment Delivery** | **Intervention delivery** | **Intervention delivery**<br>[2,12,181,182,194,195,213,253,258,266,271,272,36,273,276–284,40,285–294,46,295–304,48,305,52,109,143,153] |

| | | |
|---|---|---|
| Is there differentiation (providers only deliver the target treatment and not other treatments), treatment competency (providers maintain skills learned in training) and treatment adherence (delivery of treatment components as intended)? | Adherence to and integrity of treatment. Measures may include frequency counts, logs, records of how many times a behaviour or technique occur. Includes whether prescribed behaviours have taken place, dose of programme content delivered to and received by participant. | Closeness of the intervention delivered in the trial to the intervention in the study protocol, or the quality of intervention delivery during the trial. This includes any tailoring or adaptation of the intervention, whether pre-conceived and defined, or ad hoc, that takes place during the clinical trial.<br><br>Exemplars<br>1. "Compliance with the New Start intervention will be monitored throughout the trial via observations and regular collection of activity records to assess adherence, to understand whether the facilitators deliver the intervention in accordance with training and 'as intended'.[276]"<br>2. "Treatment fidelity pertaining to treatment delivery includes ensuring that the content and dose are consistent as well as adherence to the manual.[272]"<br>3. "In what ways, if any, did the teachers amend the programme? What were the reasons for any amendments?[277]<br>4. "Surgeons were allowed to apply their own techniques within limitations of the protocol."[253] |
| Treatment Receipt<br><br>Was the treatment delivered to the participant actually "received," or understood accurately by participants? Did participants demonstrate knowledge of, and ability to use, the skills or recommendations learned in treatment. | Intervention receipt<br><br>Elements that focus on whether participants received the treatment, and whether essential elements were provided in the treatment. | Intervention Receipt[143,272,301,305–310]<br><br>Trial participants' exposure to the intervention and its active ingredients, or dosage received, their understanding of the intervention or skills, and whether they can perform the intervention related skills *during the trial*.<br><br>"Receipt relates to skill use in the intervention setting (e.g., learning goal-setting), and enactment relates to skill use outside the intervention (e.g., planning for PA sessions).[311]"<br><br>Exemplars<br>1. "whether the treatment that was delivered to the participant was actually "received" by the participant.[306]<br>2. "Receipt of treatment focuses on exposure of the participant to the intervention and their ability to understand the skills and perform the treatment-related behaviour skills during treatment delivery.[301]"<br>3. "Dose is either defined as 'dose delivered', i.e. the number of components of the intervention delivered, or as 'dose received', i.e. the extent to which the participants used the components of the intervention as intended.[307,308]" |
| Treatment Enactment<br><br>Can participants perform treatment related behavioural skills and cognitive strategies in relevant real-life settings? Are skills implemented in appropriate situations and time to have the intended effect on clinical and research outcomes? | n/a | Participant enactment of treatment skills[2,40,44,51,52,182,272,305,312]<br>Participants' implement the skills in daily life, or real-word settings.<br><br>Exemplars<br>1. "Enactment of treatment skills assesses the participants' ability to perform the intervention skill in real-world settings?[52]<br>2. "Enactment assessment and monitoring of participant behaviour outside of the intervention.[272]"<br>3. "Treatment enactment, which has to do with whether the participant actually uses the learned strategies in day-to-day life, is more challenging to measure but could be ascertained using self-report and proxy report instruments given at some point after the trial.[44]" |

Table 2.3: NIH-BCC, Gearing's Comprehensive Intervention Fidelity Guide Fidelity, and Best-fit framework Frameworks

## BFFS Step 4: Coding of the data

The first review of fidelity data (quotations) extracted from eligible papers a resulted in 14 fidelity concepts derived from fidelity definitions, descriptions, components, or processes supporting, maintaining, or monitoring fidelity (table 2.4). These were then collapsed through thematic analysis into overarching fidelity themes, supported by exemplars (direct quotations) from included papers, and fed into the BFF framework in the steps 5 and 6.

## BFFS Steps 5 and 6: Thematic analysis and synthesis

Some fidelity themes created in step 4 were collapsed into each other and into an existing BFF category (table 2.4). For example, themes for "intervention design," "intervention differentiation," "study design," "study protocol," and "development of study manuals or guidance materials" were collapsed into the BFF's "intervention and study design" theme. This theme includes elements related to the design of both an intervention and the study assessing its effects.

Similarly, themes for "participant engagement with interventions" and "participant enactment of intervention skills" were collapsed into the BFF's "participants enactment of treatment skills (table 2.4)." Engagement has been described as an umbrella term that includes both skill enactment and intervention acceptability.[133,142,311] Reviewing quotations supporting both

enactment and engagement, enactment was often used synonymously with participant engagement to describe a range of behaviours and perspectives influencing how participants interact with therapeutic interventions.[313]

Secondary thematic analysis and synthesis of the BFF themes and supporting quotations (exemplars) was undertaken to refine the framework further. The "intervention and study design" and "interventionist training" themes were expanded based on exemplars emphasizing the need to include interventionist training strategies and skill monitoring in the design phase of a trial and monitor drift of provider skills during intervention delivery(table 2.4). Similarly, the "tailoring and adaptation" BFF theme was expanded to accommodate differentiation of allowable interventions tailoring or adaptation prespecified in a study protocol from unintended modifications made to interventions during a trial.[260,314]

After secondary thematic analysis and discussion among researcher-reviewers, fidelity data extracted from the included studies supported all but one concept in the best-fit framework. Enactment was removed from the finalised fidelity framework. Enactment was the least frequently addressed component of intervention fidelity in studies assessing the reporting of intervention fidelity in physical complex intervention literature.[171,172,181] Thirty-one of 36 papers enactment as a component of fidelity described it encompassing participants' behaviour *outside* of the clinical trial or intervention, differentiating it from receipt, or participants' use of intervention skills during the intervention itself.[39,40,226,311,315]

A trial participant may receive a treatment delivered with perfect fidelity, and yet be unwilling or unable to apply it in daily life.[2,311] This may be influenced by a variety of factors not related to the degree of fidelity with which the intervention was delivered during the trial, such as participants forgetting to do it, lacking a suitable setting, not seeing the intervention as being relevant to them, or losing interest in the intervention.[2,51] Enactment may reflect intervention acceptability, or participants' affective attitude or responses to the intervention, rather than the fidelity with which it was administered in the trial.[311,316] Enactment may relate to treatment effectiveness in influencing participants' behaviour, rather than the fidelity with which treatments were delivered during a trial.[12,40,317,318]

While it is possible to ascertain if participants or their caregivers have understood what an intervention is meant to achieve or can perform intervention activities in the trial (both aspects captured in Receipt), participant enactment during trials may not be measurable if an intervention does not involve participants learning a set of measurable skills.[319,320] Measurement of enactment may also be impractical due to difficulty defining what it constitutes and how to capture and analyse data for it.[40,226] It is also unclear how enactment differs from other concepts describing participant behaviours frequently used in rehabilitation literature, such as participant engagement or equipoise.[133,142] Consequently, treatment enactment was removed from our model of clinical trial intervention fidelity. The remaining fidelity themes formed the new fidelity framework, described in step 7.

| Fidelity concepts derived from eligible papers and number of quotations (n) | Fidelity themes created through thematic analysis and an exemplar. | Best-fit framework theme that served as a platform against which to consider the fidelity themes. | New concepts or fidelity themes created through secondary analyses and reciprocal translation | Final fidelity themes forming the new framework and its key points |
|---|---|---|---|---|
| **Intervention design (**32)<br><br>**Intervention specification (intervention details)** (27) | **Intervention design**<br><br>"Defining the active ingredients: using theory or past research to delineate the intervention's active ingredients in clear operational terms, which should be guided by beliefs explaining why they should be successful.[321]"<br><br>"Conceptualization of a theoretical framework underpinning a treatment allows the key components that underlie the intervention to be measured and reported.[181]"<br><br>"A protocol review group should ensure that the intervention reflects the theoretical model or hypothesis.[52]" | **Intervention and study design**[2,12,268–271,40,51,67,153,181,265–267]<br><br>Treatment theory and mechanism of action, identification of active ingredients, differentiation from control or comparator interventions.<br><br>Identification of elements of the design of the trial investigating the intervention, including the trial's programme model and development of study procedures and/or treatment manuals. | | **Design:** The core elements of the intervention and the protocol for the clinical trial to evaluate its efficacy or effectiveness.<br><br>Detailed description of the intervention: Describes what intervention procedures, processes, or activities providers carry out. Describes what the experimental intervention(s) looks like, includes, or how performed. |
| **Intervention differentiation: experimental vs control, comparison, or other interventions** (26) | **Intervention differentiation**<br><br>"Treatment differentiation requires that the experimental intervention and the intervention comprising the control condition differ in the intended manner.[267]"<br><br>"Finally, program differentiation refers to how the intervention being delivered is different and distinguishable from other interventions.[143]"<br><br>"treatment differentiation, that is, ensuring that the experimental intervention condition differs from a control condition (i.e., showing much higher adherence and competence to the treatment model).[194]" | **Intervention and study design** | | Design |
| **Study protocol or design of the trial** (36) | **Study design**<br><br>"Treatment fidelity practices relating to design ensure that a | **Intervention and study design** | **Interventionist training needs to be built into the design of the trial.** | **Design:** The elements of a study protocol to test the efficacy or |

| | | | | |
|---|---|---|---|---|
| | study adequately tests its hypothesis in relation to its underlying theoretical and clinical processes[306]"<br><br>"Study design should be based upon a theoretical model or hypothesis and all objective measures should reflect this.[52]"<br><br>"Fidelity practices related to study design help investigators discern whether the study will adequately achieve the aims and test the hypotheses that have been set forth.[23]" | | 46,52,174,181,278,296,297, 322<br><br>"Identify key elements of provider briefing regarding conduct of trial. Develop trainer materials, quality standards and minimum experience levels[322]"<br><br>"Identify key elements of provider briefing regarding conduct of trial. Develop trainer materials, quality standards and minimum experience levels[322]" | effectiveness of an intervention are clearly described to ensure fidelity.<br><br>Includes a provider training plan: How providers will become proficient with intervention/study procedures to support fidelity. |
| Intervention or study manuals or guidance for providers (22) | **Development of study manuals or guidance materials**<br><br>"The content, frequency, duration and quality of the intervention can be delivered as set out in the intervention delivery manual.[323]"<br><br>"To ensure the same treatment within condition a detailed treatment manual and a treatment protocol/ checklist is used for each patient separately manual has also been developed to guarantee that the treatment will be unchanged during the course of the study.[324]"<br><br>"development of a treatment manual that includes information about treatment dose (length and number of contacts) and the specific content of each contact, standardization of therapist training, monitoring of the intervention with fidelity checklists, and inclusion of strategies to measure the client's comprehension and enactment of the intervention principles addressed.[325]" | **Intervention and study design**<br><br>"The manual of procedures (MOP; described later in this article) is a carefully constructed book that details the operating procedures for the study and procedures for training personnel in the administration of outcome measures and interventions. For example, our MOPs detail the ways in which each assessment will be administered, including the point at which it is given, by whom, in what environment, and using which equipment.[23]" | | **Design:** Describes physical or informational materials or methods used to train intervention providers in study methods or intervention delivery, or to train or help participants in carrying out intervention activities. E.g.: Intervention manuals, videos or instructional aids, exercise sheets, etc. |
| Specification of trial procedures to evaluate | **Study protocol**<br><br>"Study design: was the guideline or protocol used to guide the study | **Intervention and study design** | | Design |

| | | | | |
|---|---|---|---|---|
| **intervention** (17) | published and was it clearly identified? were standardized or validated tools used to measure patient reported outcomes?[290]"<br><br>"The degree to which a therapist implements an intervention under research conditions (treatment fidelity) is dependent upon the extent and operationalization of the intervention and skill level of the therapist.[174]" | | | |
| **Intervention ist training with intervention or trial procedures** (56) | **Interventionist training**<br><br>"Training was designed to ensure satisfactory delivery of the intervention to study participants. Trainings were tailored to account for different backgrounds and past training experiences of the FHPs.[272]"<br><br>"Provider training: strategies that address preparation for uniform delivery of treatment by providers/coaches. Standardized training ensures that training is conducted similarly for all providers. Another goal of provider training is **minimizing drift in provider skills."**<br><br>"The PIPT program was designed to promote treatment fidelity by providing quality training that impacted key provider factors and that could be replicated. Thus, we incorporated quality improvement strategies and measures (physical therapist attitudes, beliefs and confidence ) to enhance treatment quality and the impact of training.[326]" | **Interventionist training**[2,12,67,153,172,213,271,273,274]<br><br>Interventionists' training and skill with delivering the intervention during the clinical trial, and their competence to deliver the intervention. | **Interventionist training and skill maintenance needed throughout trial.**[46,52,174,181,278,296,297,322]<br><br>"Fidelity of treatment delivery focuses on ensuring the intervention is delivered as intended. Many of the concerns within delivery of treatment overlap with strategies for training and study design, including controlling for provider differences[278]"<br><br>"Provider training attempts to standardize the treatment protocol and minimize its fluctuation by assessing knowledge during and post-treatment.[52]" | **Design and Delivery**<br><br>Describes provider qualifications, background, expertise, and any training given in the intervention or study procedures as part of the trial. Also includes any strategies to monitor drift in provider skills with the intervention during the trial. |
| **Intervention ist competence or characterist ics** (44) | **Provider competence**<br><br>"Intervention fidelity also requires that the appropriate background and experience level of the study therapists is identified and ensured.[174]"<br><br>"accurate delivery is highly dependent on the skill, experience, | **Interventionist training** | **Build in methods to prevent interventionist skill drift.**[46,52,174,181,278,296,297,322]<br><br>"Provider Training involves standardizing | Design and Delivery |

| | | | | |
|---|---|---|---|---|
| | and knowledge of the interventionist.[267]"<br><br>"a specific intervention cannot be delivered until those delivering it have learned to do so in a standardized way.[275]" | | training between providers and ensuring they are trained to clear criteria and monitored over time.[46]" | |
| **Intervention fidelity, Integrity, or quality of treatment delivery (103)** | **Intervention integrity (fidelity)**<br><br>"Treatment fidelity pertaining to treatment delivery includes ensuring that the content and dose are consistent as well as adherence to the manual.[272]"<br><br>"The intervention was implemented with all the planned components. ...the treatment dose was equivalent and within the stipulated range within and across conditions.[327]"<br><br>"Treatment fidelity has four components. Integrity: was the treatment delivered as intended? Another domain of treatment fidelity, integrity (how consistently the information was delivered)...[280]"<br><br>"Ensuring fidelity through treatment delivery is focused on processes that ensure the treatment is delivered as designed and focus on standardizing and improving delivery as well as assessing adherence.[305]" | **Intervention delivery**<br>2,12,181,182,194,195,213,253,258,266,271,272,36,273,276–284,40,285–294,46,295–304,48,305,52,109,143,153<br><br>Closeness of the intervention delivered in the trial to the intervention in the study protocol, or the quality of intervention delivery during the trial. This includes any tailoring or adaptation of the intervention, whether pre-conceived and defined, or ad hoc, that takes place during the clinical trial. | **Delivery (fidelity) refers to actions of the interventionist**<br><br>"Delivery: intervention is delivered as intended. It refers mainly to actions of the interventionist.[309]"<br><br>"Treatment fidelity related to delivery of treatment considers that the interventionist delivers the intervention as intended.[301]"<br><br>"The degree to which a therapist implements an intervention under research conditions (treatment fidelity).[174]" | **Delivery**<br><br>Describes strategies planned or used to monitor, maintain, or improve intervention compliance (treatments delivered or performed as intended in study protocol) or the quality of intervention delivery. May include terms such as treatment or intervention fidelity, integrity, compliance, per protocol.<br><br>Describes how these are monitored. Example: fidelity checklists, audit of session notes, video or audio recording of intervention sessions, supervision during intervention, participant logs, diaries, worksheets, etc. |
| **Tailoring and adaptation of interventions to individuals or settings (14)** | **Tailoring, adaptation, modifications**<br><br>"In what ways, if any, did the teachers amend the programme? What were the reasons for any amendments?[277]"<br><br>"The standard elements of the treatment were then tailored, such that each participant's clinical | **Intervention delivery** | **Allowable tailoring or adaptation of interventions prespecified, in protocol or vs unintended modification during trial**<br><br>"In order to accurately evaluate | **Design:**<br>If intervention was planned to be, or allowed to be, personalised, titrated or adapted during the trial: why, when, or how this was to be done. What |

| | | | | |
|---|---|---|---|---|
| | presentation (e.g., strength, pain severity, swelling) as well as the presence of co-morbidities (e.g., back and hip pain or pathology) were taken into consideration, and exercises were chosen and progressed by the physiotherapist based on each participant's response to exercise load."[328]<br><br>"Some specifications of interventions allow for local adaptation. Even if they do not explicitly do this, local adaptations may be made to improve the fit of the intervention within the local context.[2]" | | fidelity of both the interventional bundle and implementation process it will be important to distinguish between non-compliance and purposive adaptations.[260]"<br><br>"At the design level, adaptability is often essential in ensuring that interventions can fit within different contexts.[259]"<br><br>"interveners were expected to adhere to a set of theoretically grounded, overarching principles related to pressure ulcer risk when tailoring the sessions to be participant- and situation-specific.[314]" | adaptations may be made?<br><br>**Delivery:** What modifications are were made to the intervention delivered in the trial? Was the intervention modified during the trial, or were deviations from protocol reported, including changes to interventions, unintended participant cross-over between groups/interventions? |
| **Participant exposure to intervention (51)** | **Intervention receipt**<br><br>"Exposure refers to the number, length, or frequency of intervention sessions or the frequency with which intervention techniques are implemented.[143]"<br><br>"Exposure refers to the extent to which the participant is in contact with the intervention's content. Exposure is often documented as the number of intervention sessions attended and duration of each session.[63]"<br><br>Treatment fidelity of receipt of treatment focuses on exposure of the participant to the intervention and their ability to understand the skills and perform the treatment-related behaviour skills during treatment delivery<br>"Processes of treatment receipt involve monitoring and optimizing | **Intervention Receipt**[143,272,301,305–310]<br><br>Trial participants' exposure to the intervention and its active ingredients, or dosage received, their understanding of the intervention or skills, and whether they can perform the intervention related skills *during the trial*.<br><br>"Receipt relates to skill use in the intervention setting (e.g., learning goal-setting), and enactment relates to skill use outside the intervention (e.g., planning for PA sessions).[311]" | | **Receipt:** Participants' exposure to the intervention and intervention components, including participant adherence to the number of sessions or activities prescribed in the study protocol (dosage received).<br><br>**Receipt:** Includes participants' understanding of intervention skills and ability to perform intervention related skills *during* the trial. |

| | | | | |
|---|---|---|---|---|
| | participant understanding and performance of intervention skills during treatment delivery.[301]" | | | "(receipt) ... addresses whether the participant can perform a new behaviour and therefore expands beyond just exposure or delivery of the intervention to the individual.[273]" |
| Intervention doses and dosage (23) | Intervention dosage<br><br>"Dose is either defined as 'dose delivered', i.e. the number of components of the intervention delivered, or as 'dose received', i.e. the extent to which the participants used the components of the intervention as intended.[307]"<br><br>"Dosage may include (a) the number of times the interventionist addresses a target or uses a technique during a given treatment session (e.g., 30 models in 30 min), (b) how long a treatment session should last, (c) how often treatment should be delivered throughout a week or month (e.g., 1 hr/week), and (d) the total length of required intervention across time (e.g., 9 months.)[266]" | Intervention Receipt | | Delivery and Receipt<br><br>Schedule, duration, intensity, or dose: Report the number of times the intervention was delivered, or meant to be, and over what period of time. Was intervention(s) performed individually or in a group, supervised or not, and where performed (e.g.: home, community, clinic...)?<br><br>"To what extent were the participants completing the prescribed activities?[288]" |
| Participant enactment of intervention or skills (36) | Participant acceptance and uptake of interventions<br><br>"Enactment: assessment and monitoring of participant behaviour outside of the intervention.[272]"<br><br>"treatment enactment focus on ensuring that cognitive and behavioural intervention elements are applied in relevant daily life situations.[305]"<br><br>"the behavioural changes that a participant makes outside a therapy session as a result of an intervention.[51]" | Participant enactment of treatment skills[2,40,44,51,52,182,272,305,312] Participants' implement the skills in daily life, or real-word settings. Participants | Participant enactment measures what participants do outside of the trial. [2,40,44,51,52,182,272,305,312]<br><br>"Treatment enactment, which has to do with whether the participant actually uses the learned strategies in day-to-day life, is more challenging to | (Enactment removed from final framework)<br><br>"Because neither adoption nor enactment measures how well the intervention was delivered as conceived and planned, they are not included as components of intervention fidelity.[284]" |

| | | | | |
|---|---|---|---|---|
| | "The uptake of a new intervention depends on its acceptance by and acceptability to those receiving it.[2]" | | measure but could be ascertained using self-report and proxy report instruments given at some point after the trial.[44]"<br><br>"Treatment Enactment. Ensuring treatment enactment is the third component of treatment fidelity. This component refers to the behavioural changes that a participant makes outside a therapy session as a result of an intervention." | |
| **Participant engagement with intervention (18)** | **Participant engagement with interventions or responsiveness**<br><br>"In this review, the term 'participant engagement' is used as an umbrella term to encapsulate constructs of fidelity that relate to participants' engagement with intervention content. This includes whether participants understand the intervention, whether they can perform the skills required by the intervention, and whether they use these skills in daily life ('intervention enactment').[142]"<br><br>"It may evaluate how far participants fully accept the responsibilities required by an intervention, how far they perceive the intervention to be useful, and, more broadly, how responsive the environment is into which an intervention is introduced. In this sense, "enactment" may be considered a potential element of participant responsiveness.[2] | **Participant enactment of treatment skills.** | "| **Removed from final framework** |

Table 2.4: Coding of fidelity data

## BFFS Steps 7: The CONSIDER Framework

The resulting fidelity framework, the **Co**mplex I**n**terventions De**s**i**g**n, **D**elivery, **R**eceipt (CONSIDER) framework (Figure 2.4), is a multidimensional construct consisting of three main components: Design, Delivery and Receipt. These encompass most of the life cycle of a complex intervention clinical trial, from study design through the clinical trial and process evaluation, with fidelity processes playing a key role in each. Their components were specified and supported with examples and direct quotations from empirical papers identified in the systematic searches in stage 1 and used to create a CONSIDER checklist to facilitate their identification in complex intervention trial papers. The checklist is described later in this chapter. Its reliability was assessed and is described in the following chapter.

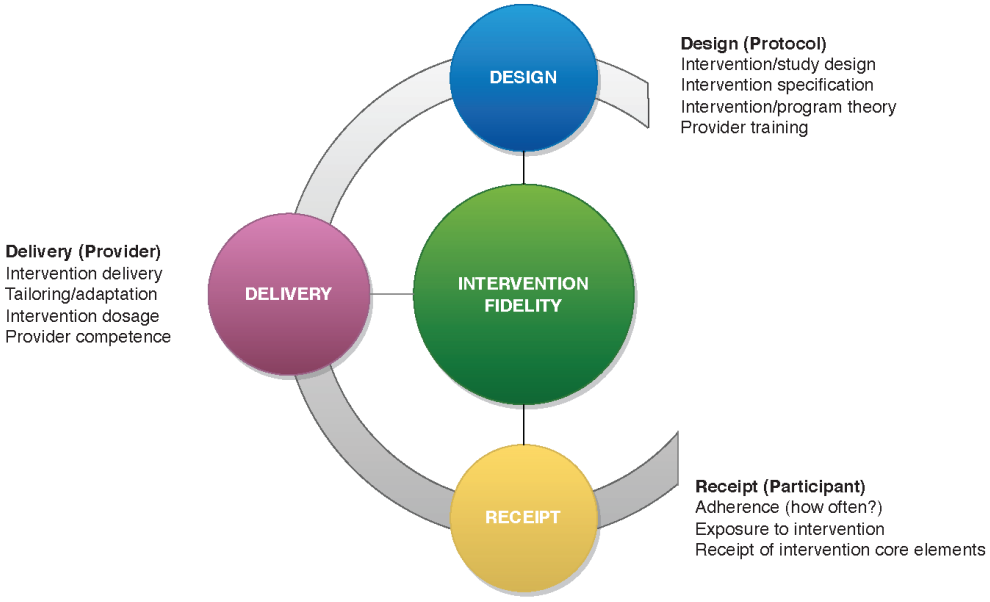



**Figure 2.4:** The Consider Framework

## The CONSIDER Framework: Design

In the design category, the essential elements of the intervention and fidelity processes in the clinical trial to evaluate it are specified.[2,43,44] It includes intervention definition, intervention or programme theory, trial design, and the provider training plan for the clinical trial.

Fifty-five papers in the thematic synthesis addressed components of intervention and study design and 34 addressed aspects of provider training. These items closely correspond to the TIDieR checklist's items; 1 (Intervention name or description), 2 "Why" (rationale, theory, or goal of the elements essential to the Intervention), 3 "What" (Materials: physical or informational materials used in the intervention, delivery or in training of intervention providers), 4. (Procedures: activities, and/or processes used in the intervention, including any enabling or support activities, 5 "Who provided" (intervention provider, expertise, and any specific training given), 6 "How" (modes of delivery of the intervention).[150] They also align with Items 11a–d in the SPRIT checklist, relating to trial protocols providing information about "each group with sufficient detail to allow replication' and 'procedures for monitoring adherence to intervention protocols.[26]"

## Elements of "Design"

**Intervention definition** is the specification of an intervention's active ingredients and components and forms the foundation for fidelity.[43,190,329,330] Active ingredients are the treatment elements hypothesized to produce intervention effects.[204] They are the essential components influencing the physiological and behavioural effects that the intervention is designed to deliver (object of treatment).[2,40,43,44,331] Examples include mechanical force applied

by a brace to a joint, lysis of intra-abdominal adhesions during laparoscopy, or learning that modifies a pattern of behaviour. Identifying interventions' active ingredients is vital for differentiation between the experimental and control interventions, ensuring that they do not overlap or provide the same therapeutic elements, confounding attribution of treatment effects to the investigated intervention.[2,42–44,195,267,280,299]

Intervention definition also describes the intervention's **treatment theory**[332], or how particular ingredients directly alter specific aspects of functioning and what actions interventionists take to deliver them to influence the object of treatment.[2,40,43,44,181,278,331] An intervention may be designed to influence the object of treatment directly or to produce effects distal to the object of treatment.[269,333,334] For example, endovascular thrombectomy (surgical intervention) with mechanical clot removal (active ingredient) may be designed to achieve vascular reperfusion (direct object of treatment). Progressive resistive exercises (active ingredient) may be prescribed to improve muscle strength as a direct object of treatment object, or as a component in a programme to improve stair climbing, a more distal outcome in which the treatment mechanism does not directly act on the outcome. Treatment theory can help ensure that interventions are targeting appropriate outcomes. More extensive use of treatment-theory has been associated with greater effectiveness and statistically significant increases in effect sizes.[335,336]

Treatment theory can help delineate interventions' key active ingredients when there are several potential ingredients present.[2,40,43,44,181,278,331] This can facilitate identification of core and flexible components of an intervention and setting of allowable parameters within which tailoring or

adaptation can take place while ensuring interventions remaining faithful to their underlying theory of action.[2,40,43,44,181,278,331] This corresponds with TIDieR[150] checklist items 9 "Tailoring" (if the intervention is planned to be personalized, titrated or adapted, then describe what, why, when, and how) and CERT[151] items 14a and 14b (tailoring and adaption, including detailed description of how exercises are tailored to the individual).

Complex interventions often require **tailoring and adaptation** of interventions to individual participants and contexts.[6,7,337] This requires identification of contextual factors and participant characteristics that may necessitate intervention adaptation.[337] Some complex interventions, such as surgical procedures, may require adaptation and tailoring that is unforeseeable during protocol planning, and strict adherence to every element of very detailed processes for these interventions may be unsafe or unfeasible. In such cases, it becomes important to determine, *a priori*, which active ingredients or components of the surgical intervention are essential for patient safety as well as intervention integrity and monitor fidelity to those while allowing necessary flexibility for the surgical provider and maintaining essential intervention fidelity.[16,338] Consideration of the limits of acceptable tailoring and pre-determining allowable parameters for in-trial intervention adaptation are also important when interventions are being evaluated against an active control or standard of care, ensuring that interventions do not deliver the same active ingredients and there is no carryover between groups, maintaining intervention differentiation.[2,40,56,273]

CONSIDER's Design category also supports key elements of trials to test the effectiveness of an

intervention. These include the trial's **programme theory** , or how interventions should be structured or administered to achieve a therapeutic outcome and goals that define the structure, process, and outcomes of a clinical trial.[339] Other key elements include best-practice methods for trial protocols and clinical trials following appropriate recommendations such as SCARE[340], PROCESS[341], CONSORT[102], SPIRIT[26], TIDieR[150], etc. These include selection of an appropriate study design, experimental and control intervention dosages, and delivery methods, etc. It also supporting the guidance of the Rehabilitation Treatment Specification System (RTSS), [43,330] and recommendations for surgical trials developed by Blencowe, et al,[169] by recommending explicit documentation of interventions' hypothesized active ingredients, treatment theory and targets of treatment in the design stage and in study protocols.

## Monitoring and reporting elements of Design

CONSIDER emphasizes pre-determining mechanisms for monitoring and reporting of intervention fidelity and ensuring interventionists can deliver the intervention competently and with high fidelity during the trial.[40,172,181,342,343] These correspond with: CERT[151] checklist item 2 (who: qualifications, expertise, or training undertaken by the exercise instructor); 5 (detailed description of how adherence to exercise is measured and reported); items 13 (when and how much, or dosage); and items 16a (describe how adherence or fidelity were measured). Accurate intervention delivery is highly dependent on the skill, experience, and knowledge of the interventionist.[40,267] Study protocols should include mechanisms to assess for ongoing supervision of intervention delivery throughout the trial to ensure delivery consistency across providers and settings, minimizing provider drift from the protocol over time.[36,44,278,290]

These mechanisms should be determined, *a priori*, and may include interventionist training in the study protocol with well-defined and study procedures,[275,344] manualisation of the intervention,[40,52,194,293] review of the manual with interventionists,[194,278] interventionist supervision, support and audit of delivery during the trial.[40,278,322] Provider training may also be enhanced by use of case scenarios and group learning experiences to help support different training needs among intervention providers.[345] Consideration of barriers to successful interventionist training such as intervention complexity, the number of treatment components and the specificity of each should also inform the interventionist training plan.[40,56,278]

Intervention manuals may contain key details about the trial design, procedures, and programme. They often include an overview of the intervention and the intervention theory, detailed descriptions or depictions of intervention activities, equipment and materials needed, mode of delivery, intervention goals and strategies, and the role and responsibilities of interventionists.[40,323] They may also provide important guidance for allowable tailoring and adaptation of interventions for individual participants and addressing problems that may arise in the intervention.[37,40,67,133] Review of intervention manuals that describe intervention and study procedures and assessments with detailed written and photographic descriptions, visual aids, exemplars or decision-making aids and can enhance provider training and intervention fidelity.[2,194,267,278,346]

## Delivery

Delivery encompasses the provision of interventions to participants as specified in the study protocol.[2,40,109] It focuses mostly on a trial's independent variables and the actions of the provider. Intervention delivery is the most frequently addressed component of fidelity in complex interventions literature.[12,67] In thematic synthesis, 60 papers described adherence to protocols, 83 addressed aspects of intervention delivery, 29 papers describing maintenance of interventions' integrity, 6 describing intervention tailoring and adaption to individual patients or providers, and 37 describing aspects of provider competence and training in the intervention to ensure fidelity to protocols or treatment integrity.

## Elements of Delivery

Key themes in "Delivery" include maintenance of intervention delivery fidelity, also often referred to as treatment integrity or procedural fidelity (was the intervention delivered as intended?) in included papers; quality of intervention delivery; adherence to treatment protocols, tailoring and adaptation of interventions within prespecified limits; ensuring providers are trained and competent to deliver interventions, and controlling for provider differences.[278] This category corresponds with TIDieR[150] checklist items 10 "Modifications" (if modified during delivery, what, why, when, and how), and items 11 and 12 "How well (intervention fidelity and adherence and assessment).[150] It also corresponds with CERT[151] item 5 (detailed description of how adherence to exercise is measured and reported), items 16a and 16b (describe how adherence or fidelity were measured, and extent to which the intervention was delivered as

planned) and The SPIRIT[26] item 11c (strategies to improve adherence to intervention protocols, and any procedures for monitoring adherence).

Aspects of Delivery may overlap with, but proceed from, components of Design. For example, adhering to intervention protocols during a trial requires that interventions be operationalized in detail in the Design phase.[40] Intervention delivery fidelity describes whether intervention components and interventionist behaviours are delivered as intended, ensuring the interventions' delivery of active ingredients.[36,190] Monitoring of intervention delivery fidelity, preserving interventions' integrity is essential for maintaining differentiation between interventions during a clinical trial. Differentiation minimizes contamination of the intervention under investigation with elements of other interventions or mixing of active ingredients between the control and experimental interventions.[293,294,297,298,345,347]

Interventionist adherence to study protocols was the most frequently identified component of fidelity of delivery in the thematic analysis, identified in 60 of the 130 papers. While adherence to study protocols supports intervention integrity, [46,48,294,296] intervention fidelity may not necessarily require strict adherence to every component of the protocol, as discussed previously in the "Design" section, and may be supported by having prespecified allowable tailoring and adaptation of interventions to individual participants or contexts in the design stage. In highly pragmatic trials, however, it may be necessary to prioritise fidelity to treatment theory or the trial's programme theory (theoretical fidelity) to maintain fidelity to interventions' underlying causal processes and reflect real-world clinical practice conditions, rather than fidelity to specific

procedures during intervention delivery (content fidelity).[58,339,348] This is explored with real-time data gathered prospectively in an ongoing pragmatic trial (ACL-SNNAP[349]) in chapter V.

An important theme identified in thematic analysis was the importance of **provider competence** (trained interventionists) for maintaining fidelity during intervention delivery, corresponding to TIDieR[150] item 5, "who provided (describe their expertise, background and any specific training given) and 3 "what (describe physical or informational materials used in training of intervention providers." Competence captures practitioners' skill in delivering the intervention and ability to comply with proscribed behaviours and avoid contaminating the intervention with prohibited components or behaviours. This may be influenced by provider training built into a trial at the design stage and providers' ability to maintain skill in delivering the intervention throughout the trial.[40,133,138] The influence of provider training and competence extends beyond intervention delivery, also encompassing non-specific treatment effects such as interventionist ability to engage with participants, warmth, and communication skills.[39,40]

Provider competence during the trial can be supported by ensuring providers are familiar with the intervention manual and can access it as a source of guidance or support, and have supervision to prevent deviation from the intervention protocol or provide guidance when unforeseen modifications are required.[40,195] Methods designed to enhance and support provider competence should control for provider differences in education, and experience with intervention delivery and monitor whether interventionists maintained the skill set learned in training throughout the clinical trial.[36,306]

### Monitoring of elements of Delivery

Delivery can be monitored by video recording and assessment of patient sessions with a fidelity checklist to ensure the intervention is delivered as specified in the protocol and intervention integrity is maintained.[109] Other assessment options include assessment of randomly sampled audio or video recorded treatment sessions,[350] observation of treatment sessions,[351,352] interviewing of participants,[353,354] provider self-assessments,[265,318] review of provider treatment notes or adherence logbooks[355,356] with comparison to the protocol or intervention manual,[40,267,357] and process evaluation to assess protocol adherence and treatment integrity.[109,276] Many of these actions can also facilitate monitoring provider competence and maintenance of skills learned in training throughout the clinical trial.[36,306]

### Receipt

While Delivery focuses mostly on the actions of intervention providers, Receipt mostly focuses on the actions of the intervention recipients.[2,52,172,358] It was most often referred to as participant adherence in the eligible papers[2,4,40,107,108,307] and partially corresponds to CERT[151] item 13, SPIRIT[26] item 11c and TIDieR[150] item 8-"when and how much (number of times the intervention was delivered, when, how much, intensity and dosage) and 11 "how well."

### Elements of Receipt

Intervention Receipt was identified in 42 papers as participants' exposure to the intervention or its active ingredients (dose), their adherence to the frequency of the intervention or attendance in interventions sessions or appointments (adherence), degree to which they followed clinic

and/or home-based components of the treatment, understanding of intervention skills, and ability to perform intervention-related skills during treatment in the trial.[36,39,327,44,133,135,138,228,273,280,301] In complex interventions literature, less focus has been placed on monitoring and reporting of intervention receipt than intervention delivery, interventionist training, or other aspects of fidelity.[40,181,226,228] Nevertheless, key components of receipt such as intervention dose, participant session attendance, comprehension and performance of intervention related skills can greatly influence intervention fidelity and intervention outcomes.[133,228]

Participants' **exposure to interventions** and their active ingredients can be represented by participants' acceptance and initiation of their allocation intervention (**participant adherence**), the frequency and intensity with which interventions are delivered (dose); the degree of participants' attendance in treatment sessions and performance of intervention activities (adherence).[359]

**Intervention dose** can be further classified as either the intervention dose delivered-the number, frequency or intensity with which intervention components are delivered by interventionists, or as the intervention 'dose received', or the extent to which the participants performed the components of the intervention or attended intervention sessions as intended.[36,39,133,228,266,273,301,327] Measures of the number of treatment sessions or units of an active ingredient participants received can be used to indicate if a treatment met its prescribed dose.[258,360] For example, dosage may be measured by the number of times the interventionist

addresses a target or uses a technique during a given treatment session, the number of times or duration with which a participant achieves a desired physiological state (e.g. amount of time spent exercising at a desired percentage of maximal heart rate during the intervention), the number of treatment sessions a participant attended or number of time a participant performed an intervention activity (for example, twice weekly over six weeks).[143,258,310,360,266,272,301,305–309] Dose can be monitored with instruments measuring participants' exposure to the intervention, such as interventionist or participant logs, intervention notes, checklists, or attendance records.[37]

Receipt also includes ensuring participants' ability to perform intervention skills during the trial.[143,258,310,360,266,272,301,305–309] This assesses not just whether (or how much) participants performed intervention activities, but also how well they did so.[152] Participants' ability to perform intervention related skills during the trial is important for supporting their exposure to the interventions' active ingredients. This is particularly important for maintaining fidelity in interventions relying on participant-generated movement, such as physiotherapeutic exercise or rehabilitation interventions.[36] Participants' ability to perform intervention-related skills may also be influenced by moderating factors such as intervention complexity and interventionists' skill in communicating with participants.[2,39,40,56] CONSIDER's emphasizes manualizing intervention components, and ensuring provider competence in delivering interventions, supporting Receipt.

### Monitoring Receipt

Receipt has been operationalized and monitored in a variety of ways in complex interventions literature.[228] These include assessment of records from intervention sessions or treatment logs,

participant attendance logs, participant-completed checklists or activity logs, field notes, website monitoring or monitoring of competition of online intervention modules, and qualitative interviews with participants. Other examples included participants being contacted by trialists or receive informational material, DVDs weblinks, emails, texts or other contacts and resources to ensure their understanding of the intervention instructions and enhance intervention receipt. Assessments of participant receipt based on attendance logs, treatment session notes, field notes, daily journals, completion of practice logs, logins/website monitoring, were generally collected during the intervention period.

## Moderating factors for intervention fidelity

The Medical Research Council's (MRC) guidance on process evaluations describes the term *context* as including, "anything external to the intervention that may act as a barrier or facilitator to its implementation, or its effects.[49]" Several such potential moderating factors for intervention fidelity were identified during the BFFS and may also need consideration when monitoring fidelity in clinical trials.[2,40] Factors outside of the intervention, such as scheduling and difficulty accessing the intervention site may influence participant receipt and engagement with interventions. Comorbid conditions reducing participants' ability to perform the intervention, or participation affected by poor interactions with the intervention or interventionist, may also reduce both intervention delivery and intervention receipt in a trial.[2,40] Providers' prior expertise with an intervention and can also influence participants' receipt and engagement and should be considered when evaluating factors influencing intervention effectiveness.[306] The acceptability of interventions to providers and provider or participant

equipoise may also influence their delivery of the intervention and participants' receipt.[310,342,359,361] While these and other factors external to the intervention lie somewhat outside of the core aims of this framework synthesis, they should also be considered as part of a trial's intervention implementation plan or process evaluation.[2,135]

## OBJECTIVE 3: Definitions of Fidelity

Ninety-five descriptions or definitions of intervention fidelity were identified in the systematic review's eligible papers. Multiple terms, such as: adherence, integrity, compliance, concordance, fidelity, or specification were used, often interchangeably, to describe concepts or processes related to intervention fidelity. Researchers conceptualised or discussed fidelity in terms of interventionists' adherence to a study protocol or treatment manual; the extent to which the intervention delivered resembled the intervention that was intended; the extent to which intervention was delivered as planned[362]; protocol adherence and acceptability; adherence and provider competence[363]; and, methodological practices used to ensure that a research study reliably and validly tests a clinical intervention (table 2.5).[268]

| Term | Example definitions |
|---|---|
| Adherence | "Although often confused with adherence, which is concerned with participants' behaviours, fidelity refers to the extent to which the study team complies with the study protocol. A participant who follows the program that his or her randomization or grouping mandates would be considered to be adhering in an occupational therapy clinical trial.[23]"<br>"The extent to which patients follow the instructions they are given for prescribed treatments.[364]" "The extent to which an individual corresponds with the quantity and quality of exercise, as prescribed by their healthcare professional.[365]" |
| Concordance | "The process of enlightened communication between the person and the healthcare professional leading to an agreed treatment and ongoing assessment of this as the optimal course"<br>"measurement of concordance of patient and provider understanding of the problem and/or treatment recommendations[290]." |
| Compliance | "Patients' obedience to recommendations with prescribed treatments.[364]"<br>"Individual therapist behaviours as compliant/not compliant with the treatment manual. This very stringent procedure showed that there were deviations from the protocol, e.g. because cues prescribed in the manual were omitted or augmented.[302]"<br>"How far those responsible for delivering an intervention actually adhere to the intervention as it is outlined by its designers.[366]" |
| Fidelity | "Fidelity: Adherence to the intervention protocol, delivery as per manual.[367]"<br>"Fidelity (degree to which the protocol was implemented as planned), the extent to which the intervention has been received by the audience.[321]"<br>Fidelity includes quality of delivery and dose.[368]<br>"Delivered as intended; adherence; integrity; quality of program delivery.[369]" |
| Implementation fidelity | "Whether a program service or intervention is being delivered as it was designed or written.[2]" |
| Intervention fidelity | "Intervention fidelity refers to the extent an experimental manipulation has been implemented as intended in a comparable manner to all participants.[258]"<br>Therapists' adherence to the intervention protocol.[370]<br>"Intervention fidelity (i.e., adherence, compliance, integrity). degree of which the intervention was delivered as originally planned.[371]"<br>"Ongoing assessment, monitoring, and enhancement of the reliability and internal validity of an intervention or treatment.[228]" |
| Treatment fidelity | "All sessions were rated as compliant adherent to manual/treatment intent.[289]"<br>"The treatment fidelity: the dose, the protocol process adherence and content of treatment delivered compared to the protocol. The extent to which the intended intervention was provided by therapists."<br>"Treatment fidelity, i.e. if the treatment was delivered as intended. [372,373] |
| Treatment integrity | "Treatment fidelity is comprised of two components. The first is treatment integrity, the extent to which interventions are implemented as intended for the duration of a study (i.e., each participant receives the intended treatment).[347]"<br>"Integrity; was the treatment delivered as intended?[279]"<br>"treatment integrity, that is, demonstrating that therapists carry out the intervention with adequate levels of adherence and competence to the treatment model or protocol.[194]"<br>"Treatment integrity typically involves two processes, adherence or fidelity to the manual, protocol, or treatment model, and competence or level of skill with which therapists deliver specified treatments.[374]<br>"Treatment integrity, or procedural fidelity... [279]" |
| Treatment quality | "Assessment of treatment quality captures the manner in which a treatment is delivered. This component of fidelity assessment seeks to differentiate between treatments implemented well versus interventions implemented poorly.[266]" |

Table 2.5: Terms and definitions for fidelity

The words most frequently used when describing or defining fidelity were fidelity (n = 260), delivery (131), training (97) adherence (93), compliance (86) protocol (84), delivered (84) intended (71) and receipt (67) (Figure 2.5).


Figure 2.5: Fidelity terms word cloud

## Fidelity and adherence

Within these definitions and descriptions, two distinct concepts emerged: Fidelity and adherence. **Fidelity** (intervention, treatment, or implementation) most often referred to the action of interventionists and the quality of their intervention delivery during the trial. Operational constructs defined fidelity in procedural terms related to the administration of a therapeutic intervention, including the integrity of treatment delivery, or the closeness or concordance of the intervention delivered to the intervention intended in the trial protocol or manual. Definitions focused on construct fidelity referred to the extent to which interventions

delivered in the trial were faithful to their underlying theoretical basis, active ingredients or clinical guidelines.[42]

The other concept, **adherence**, most often referred to action of participants, or the extent to which participants complied (compliance) with, performed, were exposed to or received the intervention in the trial.[23] For example, adherence was exemplified by participants following their randomised group allocation and not crossing over to the comparison condition in surgical or rehabilitation trials,[2,4,26,27] and attending the prescribed number or therapy sessions or fully performing the required number of home exercises or activities in rehabilitation trials.[2,4] Participants who are either underexposed to the intervention because they did not attend intervention sessions or failed to perform intervention activities, or were overexposed to the intervention because they received greater intervention dosage or greater number of intervention sessions than prescribed in the study protocol would be considered non-adherent.[23]

### Integrated definition of intervention fidelity

Through thematic analysis, an integrated definition of intervention fidelity was derived for this thesis in which intervention fidelity is an umbrella concept encompassing two distinct but related and interacting components: intervention fidelity and participant adherence. Both are essential for a clinical trial to be faithful to its protocol, and both can influence study outcomes, individually or together:[3,19–21]

5. **Intervention delivery fidelity (fidelity)** focuses mainly on the actions of the interventionist. It encompasses the quality of intervention delivery or performance within the trial, and reflects the correspondence of interventions delivered in the trial with the

intervention specified in the study protocol, or in accordance with study procedures, treatment manuals, etc.

6. Participant adherence (adherence) focuses mainly on the actions of the participant. Adherence encompasses both whether participants accept and initiate the intervention allocated, and how well they comply with the prescribed, allocated intervention. For example, this could be represented by measuring whether participants attended the required number of therapy sessions, the frequency of intervention sessions or frequency participants performed intervention activities (for example, completed a home exercise a certain number of times per week). This definition also parallels adherence as defined in pharmaceutical trials.[25]

## The CONSIDER Checklist

The components of the CONSIDER framework were operationalised to create the CONSIDER fidelity checklist (chapter III) to facilitate identification if intervention fidelity in clinical trial papers in the meta-epidemiological study of fidelity in complex intervention clinical trials in chapter IV. To further facilitate their identification, the items from the CONSIDER framework and checklist were also aligned with corresponding items on the, TIDieR[150], CERT[151], and SPIRIT[26] checklists. An elaboration and explanation document and scoring guide, supported with examples and direct quotations from empirical papers identified in the systematic searches in stage 1 was also developed and the checklists' reliability assessed. These steps are the subject of the following chapter (chapter III).

## DISCUSSION

Intervention fidelity is an essential part of conducting intervention research and implementing the findings into clinical practice. [16,37,45,67] A lack of a uniform definition of fidelity and its key components has been identified as a barrier to fidelity planning and intervention implementation in clinical trials and their translation to clinical practice.[40,375] The broad range of fidelity terms, definitions and concepts used in complex interventions literature also makes it difficult to systematically identify fidelity reporting in clinical trials. The CONSIDER framework synthesizes key aspects of intervention fidelity from 269 empirical and methodological papers to create a fidelity framework developed specifically for clinical trials of physical complex interventions.

CONSIDER was developed as a basis from which to frame and investigate intervention fidelity in this thesis, but also represents an important first step in providing practical guidance for intervention fidelity in the planning and implementation of clinical trials in domains involving physical complex interventions such as physiotherapy and rehabilitation. To the best of our knowledge, it is the first "Best-fit" framework synthesis of intervention fidelity and first empirically based fidelity framework created specifically for complex interventions in the physical domain. Further development of the framework and checklist with broader input from a wider range of stakeholders is needed to refine the framework and enhance its applicability for future evaluations of complex intervention clinical trials.

Much of the fidelity monitoring identified in this framework synthesis and in previous systematic

reviews focuses on intervention components adhering to trial protocols or participant adherence to treatment frequencies.[40,181,226,228] While these are important elements of intervention fidelity, focusing on these alone neglects the influence of other key elements of fidelity on the outcomes of a clinical trial. [2,138,169] For example, poor treatment specification or interventionist training in the Design stage may lead to suboptimal intervention delivery, which may lead to poor participant exposure to the intervention (Receipt). CONSIDER offers a more complete conceptualization of fidelity, encompassing both the interventions and the design and conduct of trials to assess their effectiveness.

CONSIDER also supports intervention fidelity that is flexible, recognizing that tailoring and adaptation of interventions may be necessary to accommodate individual participants and clinical contexts. [47,376] Rather than rigid adherence to large numbers of intervention components, CONSIDER emphasises tailoring and adaptation within pre-determined boundaries that is based on fidelity to interventions' treatment theory, retaining intervention fidelity while not impeding the application and effectiveness of complex interventions.[2,138,169]

The CONSIDER framework can be used in conjunction with existing clinical trial models or frameworks to contribute a deeper, broader conceptualization of intervention fidelity. It complements other, established design and reporting frameworks such as CONSORT, TIDieR[150] or CERT[151]. Enhancing fidelity in the design and intervention implementation of clinical trials supports enactment of processes reported on TIDieR[150], CERT[151]and SPIRIT[26]. While Intervention fidelity is a separate concept from intervention reporting, an important relationship exists

between the two. Intervention fidelity cannot be reported adequately if it has not been previously considered or monitored. The processes which support intervention fidelity also support transparency and enhance the documentation of intervention details needed to support reproducibility and the dissemination of evidence-based methods.[16,43,67,105,330,342] Variable and imprecise description of intervention components in clinical trial papers makes it difficult to identify the active ingredients interventions were meant to deliver or whether departures from the intended intervention took place.[1,43,150,197,222,224,279] CONSIDER complements the recommendations of the Rehabilitation Treatment Specification System (RTSS), emphasizing identification of interventions' active ingredients and treatment theory and the development of empirically testable interventions.[43,330]

The recommendations of the CONSIDER framework can also support the development of better comparison-control treatments for complex interventions, complementing the DITTO (**D**econstruct, **I**dentify, **T**ake out, **T**hink risk, **O**ptimise) framework[377] and ASPIRE guidelines for placebo and sham intervention controls in surgical and rehabilitation trials.[262] Placebo controls or sham interventions appear similar to the experimental treatments but lack their active ingredients,[378–381] and minimize the risk of biases such as expectation, performance, detection and confirmation biases.[54,262,378–383] These biases are high in trials of surgical and physiotherapy interventions and weaken the validity of studies' findings, but placebo interventions are methodologically difficult to construct and present to patients.[262]

Interventions' active ingredients and fidelity, as specified in CONSIDER, can be manipulated to

move them from the experimental intervention to a placebo intervention delivering no or very low dose of the active ingredients.[384] Once the intervention being evaluated has been operationalized, its placebo control can be constructed by "moving the needle" between varying levels of intervention fidelity to produce placebo interventions that are identical to the experimental surgical procedure but lack its active or essential components. For example, in the Can Shoulder Arthroscopy Work? (CSAW) trial, a RCT assessing the clinical and cost-effectiveness of arthroscopic subacromial decompression for shoulder pain, the essential surgical element (bone and soft tissue removal) was manipulated to randomise participants to an arthroscopic surgery group with the essential surgical element (active arm), a diagnostic arthroscopy only without the essential surgical element (no spur removal) placebo arm) or an active monitoring group.[9]

No similar guidance exists for the construction of placebo interventions in Physiotherapy or Physical and Rehabilitation Medicine (PRM), encompassing a larger spectrum of medicine and rehabilitation disciplines.[382] The intensive provider-participant contact and multi-modal nature of physical therapy and PRM interventions present unique challenges for the construction of placebo-controlled trials.[382] The International Placebo Symposium Working Group was convened in 2010 to address these and made a number of recommendations that would be supported by use of CONSIDER, including greater efforts to reduce variability in intervention implementation, greater evaluation of the isolated components of rehabilitation interventions, and use of structural equivalence, in which the experimental and placebo groups have similar degree of therapeutic contact.[382]

## Strengths, limitations, and future directions

CONSIDER has been developed to encompass the unique challenges and opportunities posed by interventions and clinical trials in domains involving physical complex interventions such as physiotherapy and rehabilitation. We followed a thorough, systematic best-fit framework approach and derived evidence from empirical, methodological, and theoretical literature in these complex interventions, supporting its applicability in their clinical trials. A reported limitation of many existing implementation frameworks and models is that they describe determinants and moderators of fidelity without elaborating on the relationships between them or the mechanisms linking them to implementation outcomes.[385,386] Our best-fit framework synthesis sought to overcome this limitation though the secondary thematic analysis and reciprocal translation of themes derived from eligible papers, extensive use of exemplars from the complex interventions literature base, and linking of concepts between and across CONSIDER stages.

Our search strategy was comprehensive and maximized sensitivity rather than precision. It was unrestricted by language and included both commercial and grey literature sources.[239] A broad range of search terms to ensure relevant papers were captured.[239] Although the search strategies used in this analysis were comprehensive and conducted in multiple search engines, it is possible that some papers describing intervention fidelity may have been missed. The lack of consensus on definitions and components of fidelity, the many terms used to describe it, and poor reporting of fidelity in complex intervention literature increase the risk that some eligible papers may not have been captured by our search terms. However, we employed citation

searching and extensive full-text screening to ensure that papers describing intervention fidelity with unanticipated terms were also captured.

Systematic reviews of complex interventions have found poor or completely absent reporting of fidelity monitoring or assessment across clinical trials.[45,169,290,387,388] This may reflect some degree of editorial constraint, in which word count limits and manuscript length restrictions limit reporting of some aspects of the conduct of clinical trials.[388,389] We attempted to overcome this by rigorous full-text screening of all papers for concepts or processes related to fidelity and searched for trials' protocols or registrations, reviewing them and searching for information about intervention fidelity when they were available. The framework synthesis undertaken in this review also aimed for conceptual saturation and generalizability, rather than statistical power [240,241], and study selection was purposive rather than exhaustive. The large number of papers in our best-fit framework synthesis maximized the likelihood that conceptual saturation was reached.

Even with our comprehensive searches, we found few reports of trials assessing intervention fidelity in surgical interventions, despite having worked with surgical trialists to enhance the search strategy's sensitivity and extensive efforts to identify application of fidelity principles in surgical trials. It is possible that some surgical trials may have been missed because the processes that support fidelity during clinical trials were described in terms falling outside of our search strategies. To overcome this, full text screening was undertaken for any surgical papers identified with our search strategies or citation searches to identify papers applying any fidelity principles, even if not labelled as such.

Nevertheless, previous systematic reviews and methodological papers have also identified a paucity of surgical trials monitoring fidelity principles. Beard et al. (2020) reviewed 96 papers describing surgical placebo controlled trials in the development of the ASPIRE guidelines, finding only four papers reporting elements of fidelity and seven reporting standardization of the intervention, a component of intervention fidelity for clinical trials in CONSIDER.[262] Methodological papers have also identified the unique challenges to intervention fidelity and adherence posed by surgical trials, including inherent and unpredictable variability in surgical procedures due to surgical findings, surgeon learning curve effects, and high potential for cross-over (poor adherence) between trial arms in trials comparing operative versus nonoperative therapy.[27] As a result of the poor representation of surgical papers in the BFFS, the CONSIDER framework represents the perspective of rehabilitation (i.e. physiotherapy, occupational therapy, speech-language therapy, exercise interventions, etc.) and not of surgery.

Finally, although the "best-fit" framework syntheses method is particularly suited for developing a comprehensive framework based on existing evidence, and our database of empirical and methodological papers was extensive, CONSIDER and the checklist are initial steps that need to be developed further with broader input from a wider range of stakeholders before they can be presented as a tools trialists should be using. In future stages of their development, a Delphi process will be needed to build consensus about the synthesized definition of fidelity, which fidelity components and qualities are most important, and which qualities should be given the

most weighting when developing and evaluating intervention fidelity in complex intervention trials.

Additionally, the applicability of the framework and checklist to study settings challenging for fidelity needs to be explored. Explanatory randomised trials are conducted under idealised conditions to give interventions the best chance to demonstrate an effect (efficacy).[390,391] These tightly controlled conditions can facilitate maintenance and monitoring of intervention fidelity. However, pragmatic randomised clinical trials are designed to evaluate the relative effectiveness of interventions under real-life[209] conditions, with diverse clinical populations,[210] and against usual care interventions.[211,212] Key aspects of intervention delivery may be less tightly controlled in pragmatic trials, creating challenges for assessing and maintaining intervention fidelity.[210,224,225] The applicability of the CONSIDER framework was investigated in an ongoing pragmatic trial of surgical versus rehabilitation management, the Anterior Cruciate Ligament (ACL) Surgery Necessity in Non Acute Patients (ACL SNNAP)[349] trial in Chapter V.

## CONCLUSION

This framework synthesis represents an important first step in addressing a gap in our understanding of intervention fidelity in complex interventions in the physical domain. While growing attention has been paid to fidelity when interventions are translated and implemented in clinical practice, far less research has focused on intervention fidelity during the clinical trials themselves. Guidance specifically tailored to the planning and implementation of intervention fidelity in clinical trials of these complex intervention is rare, and fidelity frameworks developed for psychology and public health trials do not translate well to physiotherapy and surgery. Failure

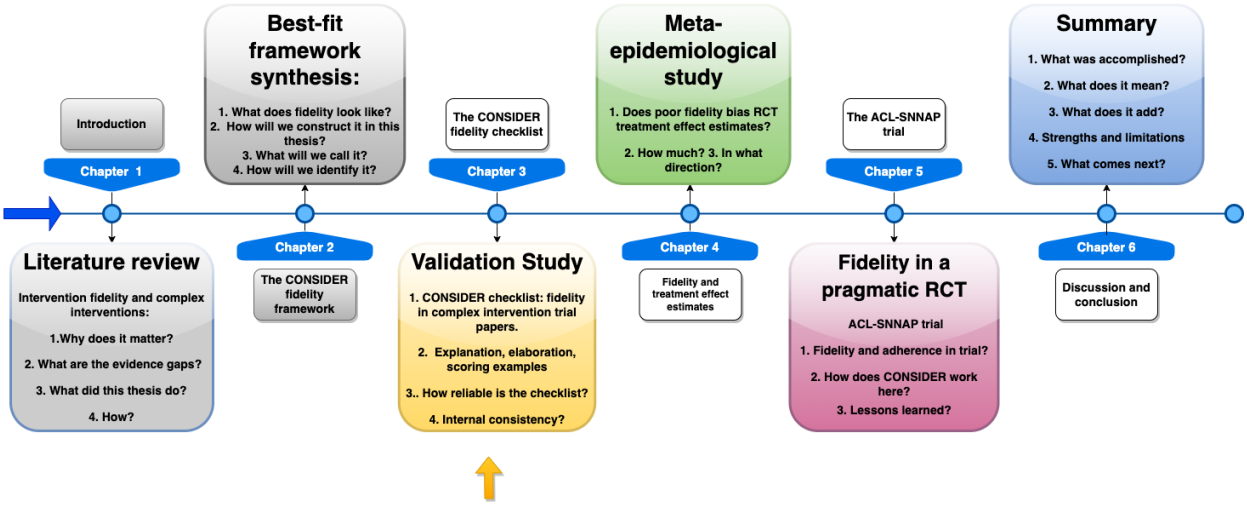
to implement interventions with a high degree of fidelity could negatively affect the accuracy and validity of clinical trials, undermining patient care and the translation of evidence-based interventions into clinical practice.[16,40,67,392]

The CONSIDER framework offers guidance for intervention fidelity in the planning and implementation of clinical trials, with implications for reproducibility and the translation of evidence-based interventions to clinical practice.[210,393–395] Further development of the CONSIDER framework with broader input from a wider range of stakeholders is needed. Ultimately, the framework may help researchers design clinical trials that enable research reproducibility and uptake, reducing waste and benefiting the practice and evidencing of complex interventions in rehabilitation.

**Chapter III:** Inter-rater agreement and reliability of The CONSIDER Fidelity Checklist

## Chapter Summary

The Complex Interventions Design, Delivery and Receipt (CONSIDER) checklist was developed to facilitate systematic and reproducible identification of intervention fidelity in complex interventions controlled trials for the meta-epidemiological study in chapter four of this thesis. This chapter described the development of the checklist and its evaluation in a cross-sectional reliability study with a purposeful sample of clinicians, researchers, trainees, and academics in complex interventions. The primary aims of this chapter were to determine the internal consistency, inter-rater reliability, and agreement of the CONSIDER checklist. Secondary aims included exploring participants' experiences using the checklist to identify items needing further development or modifications needed to improve the checklist's usability.



A comprehensive, reliable fidelity checklist was needed to facilitate systematic and reproducible identification of intervention fidelity reported in complex interventions RCTs for the meta-epidemiological study in chapter four of this thesis. To this end, the **Co**mplex I**n**terventions De**si**gn, D**e**livery and **R**eceipt (CONSIDER) checklist was developed from the Best Fit Framework Synthesis in Chapter III. This chapter describes its development and assessment of its reliability, inter-rater agreement, and usability.

## BACKGROUND

Systematic reviews of fidelity in complex interventions in education, public, and behavioural health have identified a number of fidelity assessment tools and checklists.[39,142,217,228] Fidelity assessment developed for these disciplines may not translate easily for use in trials of complex interventions in the physical domain, whose unique characteristics require fidelity assessment tools tailored to their needs.[40,45,63,219,396] For example, the NIH Behavioural Change Consortium's (NIH-BCC) comprehensive treatment fidelity checklist contains 29 items assessing fidelity across five domains: Study Design, Provider Training, Treatment Delivery, Treatment Receipt, and Treatment Enactment, with items scored as present or absent.[39] Key aspects of the NIH-BCC checklist, such as its constructs for treatment enactment and receipt reflect characteristics of psychological behavioural health interventions and do not translate well for use in trials of complex interventions in the physical domain such as rehabilitation, as described in chapter II.

Additional concerns with available checklists such as the NIH-BCC include their use of

dichotomous scoring for fidelity (items scored as presence or absent), rather than ordinal scoring indicating the degree of meeting a recommendation.[397] Determining whether fidelity was achieved or maintained requires consideration of multiple aspects of clinical trial conduct, which may be achieved to varying degrees.[49,152] Additionally, many existing fidelity assessment tools measure only one or two dimensions of fidelity, such as intervention delivery, treatment differentiation or provider training, or measure either the actions of interventionists or participants, but not both.[60,143] However, intervention fidelity is a multidimensional construct, as described in Chapter II, and optimal fidelity assessment tools should also be multidimensional.[60]

Most fidelity assessment tools are developed for use during or just after clinical trials (during process evaluations).[2,12,142,152] A number of checklists have been established to assess intervention implementation in clinical practice, but their reliability has not been broadly established.[2,12,142,152] Far less is available to guide assessment of fidelity in trial publications. Customized criteria for assessing fidelity in trial publications have been reported in individual meta-analyses of complex interventions, but these were not developed for wider use and their reliability has not been evaluated.[178,180,398]

## CONSIDER Framework and Checklist

The **Co**mplex I**n**terventions De**si**gn, D**e**livery and **R**eceipt (CONSIDER) checklist was created to facilitate reproducible, transparent identification of intervention fidelity for the meta-epidemiological study in chapter IV of this thesis (figure 3.1). It is a multidimensional checklist for assessing intervention fidelity reported in physical complex intervention trial publications. It was

developed through the Best-Fit Framework Synthesis[244] described in chapter 3, and following the "Guidance for Developers of Health Research Reporting Guidelines" by Moher, et al.,[399] Ginsberg et al.'s recommendations for fidelity assessment[152], and Bond and Drake's standardized methodology for developing and validating fidelity scales for evidence-based practices.[132]

Bond and Drake describe three stages for developing a fidelity scales: 1.defining scale content; 2. developing data collection procedures and; 3. assessing psychometric properties. These, and their components, are described below.

Step 1: Following Bond and Drake's process, the first step in "**defining scale content**" was to identify the purpose of the checklist, including its primary users. The CONSIDER checklist was developed to facilitate systematic and reproducible identification of intervention fidelity in complex interventions clinical trial publications for the meta-epidemiological study in chapter four. It was designed to facilitate assessment of intervention fidelity in trial publications, rather than ongoing assessment of fidelity during the conduct of clinical trials.

Step 2: In this step, "**developing data collection procedures**," rigorous theoretical or empirical work was undertaken to inform the content of the tool and operationally define 3–5 items assessing each scale component. This included identifying the key components of the focus of the scale and supporting them with direct evidence from the literature. This was accomplished through the systematic review and best-fit framework synthesis (BFFS) in the previous chapter. Multiple items (criteria) for fidelity concepts were used to build redundancy into assessment of fidelity components (for example, intervention delivery and receipt).[152]

Bond and Drake recommend that a fidelity scale should have no more than 25 items, reducing labour intensity for their users, and the items should be operationally defined with concrete and observable criteria. The CONSIDER checklist's 8 items were operationally defined with concrete data and supported with an elaboration, explanation, and scoring guide containing direct, illustrative quotations from empirical papers for each criterion and score. The checklist items were also aligned with corresponding items from the Template for Intervention Description and Replication TIDieR[150], CERT[151] and SPRIT[26] checklists.

Next item calibration and item scoring was developed. The CONSIDER checklist was developed from a premise that intervention fidelity is a complex, multicomponent construct depending on a number of actions and criteria during trial conduct, which may be achieved to varying degrees.[49,152] Reflecting this, item scoring on the checklist was not binary (i.e., yes/no, or present/absent), but ranged from either three to five response options, from "absent to mostly present." The scoring structure also followed the recommendations of Gearing's CIFG, which assesses intervention fidelity across 22 items that are scored as 0-absent/minimal, 1-moderate, 2-extensive.[40] In the CONSIDER checklist, items are scored as 0-absent, 1- somewhat discernible, 2-mostly discernible. The complete absence of a criterion was separated from it being minimally discernible to help distinguish between criteria that are poorly described, but present, from those that are absent altogether.

Items in CONSIDER assessing the degree of fidelity maintained or achieved (items 7 and 8) use 75% as a cut-off point. This was based on three factors. The NIH-BCC fidelity checklist developed

by Bellg et al. and adapted by Borelli et al. defines “high treatment fidelity” in studies when they scored 80% or greater on checklist items.[39] Summerfelt (2003) also calculated the theoretical impact of a loss of fidelity of greater than 25% on research studies’ statistical power and its implications for interpreting their outcomes.[53] Adherence, and its influence on trial outcomes, has also been investigated extensively in pharmaceutical research. Systematic reviews of adherence in pharmaceutical trials report 80% as the most commonly used threshold to define high adherence.[25] Given these findings and the NIH-BCC’s recommendations, achieving or reporting 75% or greater fidelity or adherence was designated as a scoring criteria for items 7 and 8 in the CONSIDER checklist.

Step 3: Finally, pilot testing (assessing psychometric properties) of the scale was performed, providing direct experience of using the scale to identify items in need of refinement and to gauge the scale’s feasibility and acceptability to users. This also allowed assessment of the scale’s reliability, validity, and usability. This chapter describes the piloting and evaluation of the CONSIDER checklist study, with the following aims:

AIMS:

I. To determine the inter-rater reliability of the CONSIDER checklist.

II. To determine the internal consistency of the CONSIDER checklist.

III. To survey users about their experience of using the checklist and its ease of use and identify areas in need of modification or further clarification.

## METHODS AND MATERIALS:

A cross-sectional reliability study was undertaken with a purposeful sample of clinicians, researchers, trainees, and academics in complex interventions. The reliability study was conducted following the guidance of Mowbray et al. in "Fidelity Criteria: Development, Measurement, and Validation[221]" and the Guidelines for Reporting Reliability and Agreement Studies (GRRAS).[400] Upon consultation with the Institutional Review Board at the author's primary institution and the revised US Department of Health and Human Services Protection of Human Subjects Guidelines[401], the feasibility and reliability study was exempt from IRB review.

### Study materials

First, a dictionary of operational definitions for intervention fidelity and an explanation and elaboration (E&E) paper with checklist items, scoring criteria, and illustrative examples from the database of papers included in the thematic analysis in chapter II (figure 3.2). The E&E paper was developed from the database of complex intervention papers and thematic synthesis conducted in chapter II. The E & E paper serves as both a manual and scoring guide for the CONSIDER checklist (figure 3.1) and includes illustrative examples from three complex intervention clinical trial papers for each score option for the checklist's 8 items.

# CONSIDER Fidelity Checklist

| Items to indicate whether fidelity supported? | Items 1-5 Scoring: 0: Absent 1: Somewhat discernible 2: Mostly discernible |
|---|---|
| **Item 1.**<br>CONSIDER Design<br>TIDieR item 3<br>CERT item 1 | **Materials**: Describes physical or informational materials or methods used to train intervention providers in study methods or intervention delivery, or to train or help participants in carrying out intervention activities. e.g.: Intervention manuals, videos or instructional aids, exercise sheets, etc. |
| **Item 2.**<br>CONSIDER Design, Delivery<br>TIDieR 4<br>CERT 8,9,13, SPIRIT 11a | **Detailed description of the intervention**: Describes what intervention procedures, processes, or activities providers carried out. Describes what the experimental intervention(s) looks like, includes, or how performed. |
| **Item 3.**<br>CONSIDER Design<br>TIDieR 9<br>CERT 14, SPIRIT 11b | **Tailoring:** If intervention was planned to be, or allowed to be, personalised, titrated or adapted during the trial: why, when, or how this was to be done. What adaptations may be made? |
| **Item 4.**<br>CONSIDER Design, Delivery<br>TIDieR 5,<br>CERT 2: Who | **Intervention providers:** Describes qualifications, background, expertise, and any training given in the intervention or study procedures.<br>**Score: 1** if only reports one of: providers' qualifications/expertise or the training provided.<br>**2** if both described. |
| **Item 5.**<br>CONSIDER Delivery, Receipt<br>TIDierR 6, 8<br>CERT 3,4,7,8, 9,12, 13<br>SPIRIT 9, 11a | **Schedule, duration, intensity, or dose:** Reports the number of times the intervention was delivered, or meant to be, and over what period of time. Was intervention(s) performed individually or in a group, supervised or not, and where performed (e.g.: home, community, clinic...)? |
| **Was fidelity monitored?** | **Individual item scoring** |
| **Item 6.**<br>CONSIDER Delivery and Receipt<br><br>TIDieR 11<br>CERT 5<br>SPIRIT 11c | **Fidelity monitoring:** Describes strategies planned or used to monitor, maintain, or improve participant adherence (attendance in intervention sessions or undertaking tasks or interventions required for the study) or intervention delivery (or performed correctly) as intended in the study protocol or manual. May include terms such as: fidelity, integrity, adherence, compliance, per protocol.<br><br>May describe how these were monitored. Examples: fidelity checklists, audit of session notes, video or audio recording of intervention sessions, supervision during intervention, participant logs, diaries, worksheets, etc.<br><br>**Score: 0** if no monitoring (as above) is reported.<br>**1** if only participant adherence monitored (e.g.: frequency of attendance or exercise, logs, dropout rate).<br>**2** if only intervention fidelity (intervention delivered as per study protocol or procedures) is monitored.<br>**3** If both adherence and intervention fidelity (as defined above) are monitored. |
| **Was fidelity reported?** | **Individual item scoring** |
| **Item 7.**<br>CONSIDER Delivery<br><br>TIDieR 10<br>CERT | **Modifications:** Was the intervention modified during the study, or were deviations from protocol reported, including changes to interventions, unintended participant cross-over between groups/interventions?<br>Score: **(-2)** Unplanned deviations from the protocol are reported, including changed treatment frequency, or cross over between intervention/control groups, OR less than 75% fidelity achieved (see item 8)<br>**0** if paper does not report/discuss modifications that took place during trial interventions.<br>**1** if tailoring or modifications made, were consistent with the study protocol or intervention manual, or 75%-90% fidelity reported (or can be calculated) if assessed.<br>**2** if authors report no, or very few, unintended intervention modifications or deviations from protocol were made. |
| **Item 8.**<br>**Was fidelity maintained?**<br>CONSIDER Delivery, Receipt<br><br>TIDieR 12<br>CERT 5, 16 | **Fidelity assessment:** If the paper reports that participant adherence or intervention fidelity were assessed or monitored: how much, or to what extent, where these achieved? Paper reports:<br>**Score: (-1)** If participant adherence or intervention fidelity were measured and either poor (<75%) or not achieved, or (-2) was scored for Item 7.<br>**0** if no report of adherence measures or fidelity assessment or results are found in the paper (Impossible to determine).<br>**1** if only participants adherence reported, > 75%<br>**2** if intervention fidelity (interventions delivered as intended is achieved (>75%), but adherence is not reported in the paper, cannot be calculated, or was not achieved (75% or more).<br>**3** if both adherence and intervention fidelity are reported to be achieved (75% or more). |

**Scoring:**

| | |
|---|---|
| **Fidelity Supported?** | Total of items 1-6: Max score 13<br>**Score 0:** "No/poor." Total of 0-5.<br>**Score 1:** "Moderate." Total of 6-8.<br>**Score 2:** "Fidelity supported." Total ≥ 9. |
| **Fidelity Monitored?** | **Score 0:** "No." Scored 0 for items 6 or 7.<br>**Score 1:** "Minimal." Scored 1 on item 6 or any score greater than 0 on item 7.<br>**Score 2: "**Monitoring present." Scored ≥ 2 on items 6 |
| **Fidelity maintained?** | **Score 0:** Absent/poor**.** Scored ≤ 0 on item 8, or -2 item 7.<br>**Score 1:** Minimal. Total score of 1 in item 8, ≥ 0 item 7.<br>**Score 2** Fidelity Maintained**:** Scored a 2 on item 8.<br>**Score 3:** Fidelity and adherence: Scored a 3 on item 8. |

**Figure 3.1:** CONSIDER checklist

## CONSIDER Checklist: Elaboration, explanation, and scoring guide

| Item | Description |
|---|---|
| 1. | **Materials**: Describes physical or informational materials or methods used to train intervention providers in study methods or intervention delivery, or to train or help participants in carrying out intervention activities. E.g.: Intervention manuals, videos or instructional aids, exercise sheets, etc.<br><br>**Score of 1:**<br>**Example:** "The treatment programme consisted of the regular model, manualized for research purposes."[402]<br>**Example:** "All 3 groups were provided with pedometers and log books to be completed."[403]<br><br>**Explanation:** The first example does not give any detail about what was in the manual, or whom it was given to. The second does not describe what the logbooks contained or recorded. Neither does it describe if the participants were trained to use the pedometers, or what kind they were ("phone app, worn at the wrist or hip…"[403]). Key details are only partially discernible from the descriptions.<br><br>**Score of 2:**<br>**Example:** "Photographic details of the exercise program were distributed to each participant together with a log book to record the number of days the exercises were performed per week."[404]<br>**Example:** "Participants also received written instructions and pictures of the stretching techniques. A daily log sheet was issued to monitor compliance." [405]<br>**Example:** "A manual was developed to guide the SHELLS intervention. The manual included information on the theory of change for SHELLS, the process of making books, the SAE strategies, and the use of the SAE strategies during a home visit. The manual was used to support home visitor training in the SHELLS intervention."[406]<br><br>**Explanation:** These items give detail about what the materials included, looked like, or were for. The third example indicates how the manual was used for interventionist training. |
| 2. | **Detailed description of the intervention**: Describes what intervention procedures, processes, or activities providers carried out. Describes what the experimental intervention(s) looks like, includes, or how performed.<br><br>**Score of 1:**<br>**Example:** "Range of motion, muscle strengthening, and endurance exercises form the base of each class. In this study, classes ranged from 45 to 60 min…"[407]<br>**Example:** "Participants were randomly assigned to one of the three intervention groups: a behavioural intervention and an educational pamphlet on the benefits of walking…. a 12-month supervised community-based aerobic walking program (SCAWP)."[403]<br><br>**Explanation:** Neither of these give sufficient details of the intervention to allow another researcher to reconstruct them. In example 1, the description is very general, and does not give information about the exercises themselves, or how much of the session each type of exercise made up. In the second example, key questions arise such as: What is the behavioural intervention? How long, far, or fast is the walking program? It would be impossible to reproduce the interventions or identify key components.<br><br>**Score of 2:**<br>**Example**: "One group performed the cross-body stretch alone by passively pulling the humerus across the body into horizontal adduction with the opposite arm, without concern for scapular stabilization (Figure 4). Each patient performed 5 repetitions of the cross-body stretch, holding each for 30 seconds."[405]<br>**Example:** "Functional strengthening exercises were performed in a circuit and were organized as a row of four exercise stations. The exercises included step-ups, chair squat, standing hip extension, and knee mid-flexion to end-range extension (in sitting position), utilizing body weight as resistance." [408]<br><br>**Explanation:** The description of the interventions in both passages makes it possible for it to be re-created by another practitioner. The papers also describe how participants should be positioned for exercises, or specific ones to be performed. |
| 3. | **Intervention tailoring:** If intervention was planned to be, or allowed to be, personalised, titrated or adapted during the trial: why, when, or how this was to be done? What adaptations may be made?<br><br>Score of 1:<br>**Example:** "this study instruction in dialogic reading was modified for community-based implementation."[409]<br>**Example:** "The home-exercise prescription was standardized for use across the trial centres, yet still allowing for minor case by case variations."[410] |

| | |
|---|---|
| | **Explanation:** In both examples, the papers report interventions were modified/adapted for use in another setting, or in response to case-by-case variations, but do not give any detail about how that was done, or what kind of variations were eligible.<br><br>Score 2:<br>**Example:** "The standard elements of the treatment were then tailored, such that each participant's clinical presentation (e.g., strength, pain severity, swelling) as well as the presence of co-morbidities (e.g., back and hip pain or pathology) were taken into consideration, and exercises were chosen and progressed by the physiotherapist based on each participant's response to exercise load."[328]<br>**Example:** "The exercise program was designed, delivered, and supervised by a physiotherapist and an exercise specialist, based on frequency, intensity, type, and time (FITT) training principles. The exercise intensity varied according to individualized functional capacity and increased progressively every week according to the ease of performance."[411]<br><br>**Explanation:** Both examples give a detailed description about how the intervention was or planned to be/allowed to be modified during the trial, and on what criteria it was based on. If a study protocol is available, it may be necessary to search the protocol for information related to tailoring, or limits of adaptation that was planned, *a priori.* |
| 4. | **Intervention providers:** Describes qualifications, background, expertise, and any training given in the intervention or study procedures.<br>Score: 1 if only reports one of: providers' qualifications/expertise or the training provided.<br>2 if both described.<br><br>**Score of 1:**<br>**Example:** "These sessions were supervised by fitness leaders in collaboration with an exercise physiologist."[412]<br><br>**Explanation:** This example tells the reader that an exercise physiologist participated, but not whether or how the physiologist was received additional training in the intervention or procedures specifically for the trial they participated in.<br><br>**Score of 2:**<br>**Example:** "Nine physiotherapists in private practices delivered both interventions. They had an average of 12 years (range 2–30 years) of clinical experience with musculoskeletal disorders. Three (30%) of these physiotherapists had postgraduate master's degree–level qualifications. All of the physiotherapists attended a 3-hour training session and were given a treatment manual."[413]<br>**Example:** "These home visitors were native Spanish speakers and had at least 3 years of experience working with children and families. One of the home visitors had a master's degree, and the other home visitor was a former Head Start parent and high school graduate. SHELLS. SHELLS home visitors were trained in how to implement the SHELLS intervention in a 2-day training session led by one of the SHELLS developers. The SHELLS training manual guided the training activities. Videotaped examples of the intervention implementation were reviewed, and role-playing episodes were practiced." [406]<br><br>**Explanation:** In these 2 examples, both the interventionists specialty or discipline (exercise physiologist, physiotherapist with 2-30 years-experience) and the training they received as part of the clinical trial (3-hour training session and manual, training in exercise prescription and counselling/supervision) are described. |
| 5. | **Schedule, duration, intensity, location, or dose:** Reports the number of times the intervention was delivered, or meant to be, and over what period of time. Was intervention(s) performed individually or in a group, supervised or not, and where was it to be performed (e.g.: home, community, clinic...)?<br><br>Score of 1:<br><br>**Example:** "a 12-month supervised community-based aerobic walking program."<br>**Example:** "Range of motion, muscle strengthening, and endurance exercises form the base of each class. In this study, classes ranged from 45 to 60 min and were taught two to five times per week depending on the pool location."[407]<br><br>**Explanation:** The first example's description does not provide key details about the frequency, duration, or intensity of the experimental intervention. In the second example, the description is imprecise. The frequency and duration are given as a range, but the range is large (2-5 times per week), making it difficult to discern the actual frequency participants achieved. It is also not possible to determine if the classes were individual or group, supervised or un-supervised.<br><br>**Score of 2:** |

| | |
|---|---|
| | **Example:** "The training program for the EXE group consisted of 150 min/wk in 2 supervised sessions of progressive mixed (aerobic and resistance) training. Exercise load for each equipment was calculated to achieve prescribed exercise intensity, expressed as percentage of maximal oxygen consumption (V˙ O2max), by the use of standard equations."[414]<br>**Example:** "For exercise at home, patients were encouraged to undertake 30 minutes of moderate intensity exercise, combining cardiovascular and resistance training, 5 days a week. Patients were given a resistance band and a pedometer to self-monitor steps count."[411]<br><br>**Explanation:** Both examples provide the frequency and intensity of the intervention, where they took place, or if they were supervised or not. |
| **Was fidelity monitored?** | |
| **6.*** <br>Delivery | **Fidelity monitoring:** Describes strategies planned or used to monitor, maintain, or improve participant adherence (attendance in intervention sessions or undertaking tasks or interventions required for the study) or intervention delivery (or performed correctly) as intended in the study protocol or manual. May include terms such as: fidelity, integrity, adherence, compliance, per protocol.<br><br>May also describe by whom or how. E.g.: investigators, supervisor, or therapist audit of session notes, video of intervention, audio recordings, supervision during intervention, participant logs, fidelity checklists, etc.<br><br>**Score:** 0 if no monitoring (as above) reported.<br>1 if only participant adherence monitored (e.g.: frequency, logs, drop out).<br>2 if only intervention fidelity (intervention delivered as per study protocol or procedures) is monitored.<br>3 If both participant adherence and intervention fidelity (as defined above) are monitored.<br><br>**Score of 1:**<br><br>**Example:** "The speech and language therapists delivering this intervention completed a questionnaire at the end of the therapy phase. The aims of this questionnaire were to quantify the hours of therapy delivered"[415]<br>**Example:** "Adherence was assessed by the number of physiotherapy sessions attended and by the number of home exercise sessions completed, as recorded by patients in a logbook. The percent home exercise adherence was calculated by dividing the number of sessions completed by the maximum required number of 48."[413]<br>**Example: "**Parents were telephoned weekly to maintain contact and to remind them to read with their child." Huebner<br><br>**Explanation:** In both cases, frequency counts (logs or questionnaires) are used to monitor the amount of intervention delivered or performed. Other examples could include attendance logs or sign-in sheets. These papers do not monitor what was done during sessions (how intense the physical activity, or how similar to the demonstrated speech therapy). The third example describes a strategy used to enhance or ensure adherence.<br><br>**Score 2:**<br>**Example**: "To address fidelity, videotaped observations of home visits were rated using the Home Visitor Facilitation of Parent–Child Interaction, Parent Engagement, and Child Engagement subscales of the Home Visit Rating Scales. An additional SHELLS Implementation subscale was developed specifically for the SHELLS project to measure the home visitor's fidelity to SHELLS strategies."[406]<br>**Example:** "A qualified exercise specialist ensured that all exercises were being performed safely and within the prescribed intensity range. In the first 2 weeks, the exercise specialist supervised all three sessions..." "Week 1 of the program was an introductory week so that patients could learn and practice each exercise while being supervised by the exercise specialist."[416]<br><br>**Explanation:** These examples demonstrate monitoring of the quality of the intervention being delivered, or whether the intervention was being performed as was intended in the study protocol, and as the interventionists were trained to deliver (first example). In the second example, the researchers are assessing whether the participants can perform the intervention activities correctly, and as intended in the study protocol.<br><br>Score of 3:<br>**Example**: "A weekly therapies log was collected... parents were asked to practice PRT daily and video record at least 10 min each week for review during group.... All intervention sessions were videotaped and at least two sessions per child were chosen at random and coded using a fidelity checklist. At least 80% correct on each of six PRT techniques was required to meet fidelity criteria."[417]<br><br>**Explanation:** The paper describes how attendance, or adherence, was monitored (therapy log), as well as how fidelity (quality of intervention delivery or performance) was monitored. |

| Was fidelity reported? | |
|---|---|
| 7.<br>Delivery | **Modifications:** Was the intervention modified during the course of the study, or were deviations from protocol reported, including changes to interventions, unintended participant cross-over between groups/interventions?<br><br>**Score:** (-2) Unplanned deviations from protocol are reported, including changes to treatment frequency, or cross-over between intervention/control groups by providers or participants, or less than 75% fidelity if assessed (see item 8)<br>0 if paper does not report/discuss modifications that took place during in-trial interventions.<br>1 if tailoring or modifications were made, they were consistent with the study protocol or intervention manual, or 75%-90% fidelity reported if assessed.<br>2 if authors report no, or very few, unintended intervention modifications or deviations from protocol were made, or >90% fidelity (if assessed).<br><br>**Score of (-2):**<br>**Example:** "No parent met fidelity of implementation at baseline. At week 12, 21 of 25 parents in PRTG, and none in the PEG, met fidelity of PRT implementation."[417]<br>**Example:** "There were differences in how the treatments were delivered due to time constraints for the nursery-based group. Activities 4 and 6 were not carried out as part of these group therapy sessions and the groups differed in how activity 3 was delivered in terms of therapy approaches .... substantial variability in the frequency that teachers followed the reading group schedule that was conveyed to them as part of the training."[418]<br>**Example:** "...participants in the control group [low-level daily physical activity] had a high level of activity throughout the study, and many exercised using HIIT [the high intensity interval training intervention]. This might have affected the study's ability to detect statistically significant differences between groups"[419]<br>**Example:** "Despite requests not to change their anti-diabetic medications unless they suffered hypoglycaemia, significant changes to medication (increase or decrease by >10% of insulin dose or 1/3rd of oral anti-diabetic medications) occurred in seven subjects."[412]<br><br>**Explanation:** In the first example, fidelity was not achieved for two of the three treatment groups. This means that participants were not performing the intervention as they were trained to, or instructed to, so that the intervention performed was different than what was planned, and the modifications departed from the intervention protocol.<br>In the second example, some of the required activities were not performed, and others were performed differently (departing from the protocol) in some groups.<br>In the third example, there was crossing-over and mixing of participants between the control and experimental interventions. Participants in the control group were performing the experimental intervention, though they were not meant to. This creates poor fidelity.<br>The fourth example reveals a significant, unexpected departure from the study protocol.<br><br>**Score of 1:**<br>**Example:** "Subjects in the home exercise group were allowed to ride a stationary bicycle if they stated that riding a bicycle was currently part of their exercise routine or if they could not walk for safety reasons."[420]<br>**Explanation:** In this example, pre-allowed adaptation to the intervention, allowing use of a stationary bike if participants were unable to walk safely, was consistent with the study protocol.<br><br>**Score of 2:**<br>**Example:** "Consultants (interventionists) completed 610 visits and submitted 138 (23%) videos and write-ups (of visits) for review; each consultant was rated for at least 22 submissions. Nearly, all (97%) submissions reached criteria."[421]<br>**Example**: "Mean results from the HOVRS provided evidence of effective implementation. These results indicate that the curriculum was being implemented as intended, the home visitors were responsively and effectively facilitating parent–child interaction, and parents and their young children were highly engaged in the home visit activities."[406]<br><br>**Explanation:** In this example, the investigators monitored intervention delivery throughout the trial. Here, they report that the intervention was delivered according to the criteria set out in the study protocol and intervention manual to a very high degree. |
| 8.<br>Delivery, Receipt<br>**Was fidelity achieved?** | **Fidelity assessment:** If the paper reports that participant adherence or intervention fidelity were assessed or monitored: how much, or to what extent, where these achieved?<br><br>**Score:** -1 If participant adherence or intervention fidelity were measured and either are poor (<75%) or not achieved, or (-2) was scored for Item 7.<br>0 if no report of adherence measures or fidelity assessment or results are Included in the paper (impossible to determine).<br>1 if only participants adherence reported, > 75% |

2 if intervention fidelity (interventions delivered as intended is achieved (>75%), but adherence is not reported or achieved.
3 if both adherence and intervention fidelity are reported to be achieved (75% or more).

**Score of (-1):**
**Example:** "Mean (s.d.) adherence to the RT program was 71±22% with 63% of the intervention group performing at least two-thirds (67%) of the sessions." "... the lower adherence to RT in our study compared with laboratory-based studies may ..."[416]
**Example:** In item 7, the third example shows cross-over of participants between the experimental intervention and control group activities, and scores a (-2). This paper would also score a (-1) here for item 8.

Explanation: These examples show poor levels of adherence and fidelity in the first example. The second example was scored (-2) in item 7.

**Score of 1:**
**Example:** "The median percentage of home exercise sessions completed was 82% (IQR 31%) by the NEXA group and 91% (IQR 26%) by the QS group."[413]
**Example:** "The majority of the children in both the Intensive and the Nursery-based groups attended all sessions. Two of the eight children in the Intensive group missed one session, and one child in the Nursery-based group missed two sessions."[415]

**Explanation:** Both examples show that more than 74% of intervention sessions were attended or completed.

**Score of 3**
**Example:** "For 89% of the families, all three sessions were held." "Results from applying this [fidelity] checklist to 77 intervention sessions (at least 2 intervention topics were selected at random for each child) revealed that 88.3% showed fidelity scores above 80% (M= 89.6%; SD = 9.0)."[422]
**Example:** "Mean results from the HOVRS provided evidence of effective implementation. These results indicate that the curriculum was being implemented as intended, the home visitors were responsively and effectively facilitating parent–child interaction, and parents and their young children were highly engaged in the home visit activities."[406] [Paper reports adherence maintained as well]

**Explanation:** In both examples, adherence was maintained (sessions were attended or activities performed sufficient number of times), and fidelity monitoring indicated that the interventions were implemented with high levels of fidelity.

**Figure 3.2:** CONSIDER Explanation, elaboration and scoring guide

## Clinical trials selection

Given that the CONSIDER checklist will be utilized to assess fidelity and fidelity reporting in a variety of studies, interventions, and study populations across rehabilitation disciplines, three clinical trial papers were randomly selected from the database of eligible papers found in the systematic review in chapter II using the randmomizr package in STATA 16.1. Three papers were chosen, rather than just one, to minimise the risk of the randomly selected paper being one with little or no information about fidelity for checklist-users to assess and to increase opportunities

to assess the performance of the checklist with a greater number of descriptions and degrees of completeness of fidelity reporting. The first block of three papers was chosen for the reliability study. These papers were two randomised controlled trials of interventions in speech-language therapy[423,424], and one of physiotherapy for knee osteoarthritis[425]. These were first assessed with the CONSIDER checklist by the thesis author (table 3.1).

| **CONSIDER item**<br><br>**RCT** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9. Was Fidelity Supported?** (Add 1-6) |
|---|---|---|---|---|---|---|---|---|---|
| 1. Pile, 2010 | 2 | 2 | 1 | 2 | 2 | 3 | 0 | 3 | 1 |
| 2. Tyler, 2011 | 1 | 2 | 1 | 2 | 2 | 2 | 1 | 2 | 1 |
| 3. Pollard, 2018 | 0 | 2 | 2 | 0 | 2 | 0 | 0 | 0 | 1 |

**Table 3.1:** CONSIDER, reference assessment

## Survey about the checklist

Participants also received a 7-item survey (figure 3.3) designed to gather information about their experiences using the checklist and identify checklist items that may need further elaboration. Descriptive statistics were used to quantify the results of participants' surveys.

CONSIDER Survey

1. How would you describe your current profession or role?

| | | | | |
|---|---|---|---|---|
| 1: Clinician | 2. Researcher | 3. Academic | 4. Any combination of 1-3 | 5. Other (please describe below) |

Other:

2. What is your level of comfort with critical appraisal or review of scholarly publications (journal articles)?

| | | | | |
|---|---|---|---|---|
| 1: None | 2. A little | 3. Moderately | 4. Mostly | 5. Very comfortable |

3. Before reading the CONSIDER overview using the tool, how familiar were you with intervention fidelity?

| | | | | |
|---|---|---|---|---|
| 1: None | 2. A little familiar | 3. Moderately familiar | 4. Mostly familiar | 5. Very familiar |

4. After reading the CONSIDER overview and appraising paper with the tool, how familiar are you with intervention fidelity?

| | | | |
|---|---|---|---|
| 1: Same | 2. A little more familiar | 3. Moderately more familiar | 4. Much more familiar |

5. With your **first** paper, how difficult did you find it using the CONSIDER tool to appraise fidelity?

| | | | |
|---|---|---|---|
| 1. Not difficult | 2. A little difficult | 3. Moderately difficult | 4. Very difficult |

6. By your **third** paper, how difficult did you find it using the CONSIDER tool to appraise fidelity?

| | | | |
|---|---|---|---|
| 1. Not difficult | 2. A little difficult | 3. Moderately difficult | 4. Very difficult |

6. How likely are you to consider a study's intervention fidelity or adherence when reading a trial publication or journal article after this experience?

| | | | |
|---|---|---|---|
| 1: Not likely | 2. A little likely | 3. Moderately likely | 4. Very likely |

7. Did you find any items particularly easy or difficult to appraise, based on the descriptions in the CONSIDER tool?

**Figure 3.3:** CONSIDER survey

## Participant (rater) sample size

The number of participants needed to assess the three selected papers with the CONSIDER checklist was determined using a sample size calculation based on the sample size formula for testing and estimating alpha coefficients derived by Bonett.[426] The coefficient of Cronbach's alpha in the null and alternative hypothesis were set to be equal to 0.0 and 0.7, respectively. Based on an alpha value of 0.05, the minimum sample size required to achieve a power of 80.0%

was 14. Twenty participants were invited to take part in the study to ensure recruitment of the minimum sample.

## Selection of participants

Participants were chosen through purposive sampling of complex intervention healthcare practitioners, researchers, and academics in the author's network of contacts. A heterogenous group of participants was invited to participate, representing the likely future users of the CONSIDER checklist. Potential participants included members of the IDEAL Collaboration (an international research group for the development and evaluation of complex interventions in healthcare), and clinicians and researchers in the Bouvé College of Health Sciences of Northeastern University, Boston, Massachusetts, the Department of Health, Kinesiology and Applied Physiology of Concordia University, Montréal, Canada, and the Harvard Medical School, Boston Massachusetts.

Potential participants had to meet the following criteria:

1. Have at least 5 years of clinical practice in a complex intervention profession in healthcare (rehabilitation, physical activity or exercise interventions, surgery, or physical medicine and rehabilitation-physiatry) OR
2. At least 5 years of experience in health professions education in one of the professions in criterion 1.
3. Have critically appraised or reviewed at least 5 scholarly papers in the previous 5 years.

## Reliability Study

Participants agreeing to take part in the reliability study received an email describing the purpose, required actions, and timeline of the study and re-affirming their ability to withdraw participation at any time. The email contained the following attachments and instructions:

1. An overview of the project.
2. The CONSIDER Fidelity checklist: you will use this to appraise three papers for fidelity.
3. A scoring examples and elaboration paper that will help you use the checklist to appraise the papers.
4. A scoring sheet for the papers (I suggest scoring them in the order they appear in the score sheet).
5. Three papers to appraise. These were randomly selected from the database of trial publications for surgery, physiotherapy, rehabilitation therapies, speech therapy, physical activity, and exercise interventions.
6. A brief survey to complete after you have appraised the three papers.

Participants were instructed to read the reliability study description and key terms first, followed by the checklist, then the E&E paper before reading and assessing the three studies with the CONSIDER checklist. Participants then independently assessed the three papers with the checklist and entered the results on a customized scoresheet (table 3.3). The Items on the checklist were scored ordinally (figure 3.1). For example, items 1-5 could be scored as 0 (absent),

1 (somewhat discernible), 2 (fully discernible). Once they completed the three assessments and survey, both the score sheet and survey were returned to the researcher (AP) by email.

Participants did not receive any prior training in fidelity assessment and were not familiar with the CONSIDER checklist. They did not receive any additional training in using the checklist and were not able to ask the thesis author any questions about the checklist or how to administer it. They were not aware that their scoring would be compared to others'.

## Data analysis

The checklist's **reliability and internal consistency,** or the degree to which items on the checklist are correlated with one another, were evaluated by calculating Cronbach's $\alpha$ in Microsoft Excel.[427] The responses in the CONSIDER score sheet were entered into an Excel spreadsheet. First, a two-factor ANOVA without replication was calculated for the checklist item scores. From these results, Cronbach's alpha was then calculated as 1-(Mean squared error / Mean squared row) and interpreted according to Cronbach's criteria.[428]

**Inter-rater reliability** was evaluated by calculating percentage of agreement[400] in participants' ratings of fidelity items across items on the CONSIDER checklist, Cronbach's alpha ($\alpha$), and Cohen's kappa[429] statistics. Percentage agreement and coefficients were based on the mean of ratings and scores for each item and paper from all participants. Percentage agreement represents how often participants assigned the same score to CONSIDER checklist items across the three papers. The percentage agreement was calculated in Microsoft Excel as the number of

matching ratings divided by the number of applicable items for each article:

The number of agreements (with the reference results)/(agreements + disagreements) x 100%

Results could vary from 0.01 to 1.0, with higher numbers indicating greater agreement. A percentage agreement of equal or greater than 80% was considered adequate.[430] Standard deviations and variance were calculated for each CONSIDER item. The mean percentage agreement for each CONSIDER checklist, mean percentage agreement and standard deviation across all raters for each paper and across all three papers was calculated. The mean standard deviation for mean ratings across all items and all three papers was calculated as the square root of the average variance across all items.[431]

Interrater agreement and reliability were also assessed with Cohen's kappa statistic, a chance-corrected measure of agreement between raters, using the "kap" command (more than two ratings, constant number of unique raters) in STATA 16.1.[429] The kappa statistic was interpreted according to Cohen's criteria (table 3.2).[429]

| Cronbach's alpha | rating | Cohen's Kappa | agreement |
|---|---|---|---|
| .01-.60 | Unacceptable | ≤ 0 as | none |
| .61-.70 | Fair | 0.01–0.20 | none to slight |
| .71-.80 | Acceptable | 0.21–0.40 | fair |
| .81-.90 | Good | 0.41– 0.60 | moderate |
| .91-1.00 | Excellent | 0.61–0.80 | substantial |
| | | 0.81–1.00 | excellent |

Table 3.2: Agreement and interrater reliability ratings

## RESULTS

Sixteen participants agreed to assess 3 papers with the CONSIDER checklist. These included 4 physiotherapists, 2 speech-language pathologists, 2 surgical fellows, 2 researchers with a background in rehabilitation, 2 academics with a background in rehabilitation sciences, 2 doctoral physiotherapy doctoral students and 2 researchers in complex interventions. One participant was unable to complete the fidelity assessments due to unforeseen circumstances. Of the 15 participants who completed the assessment, 4 participants were in Canada, 4 in the United Kingdom, and 7 in the United States. Assessment of papers began in September 2021 and ended in November 2021.

### Reliability and internal consistency

Inter-rater agreement for the 15 participants was high, with a Cohen's kappa of 0.861 (table 3.3). The CONSIDER checklist had good internal consistency, with a total Cronbach's $\alpha$ of 0.82. The 15 raters also achieved a high percentage agreement with the reference ratings across the checklist's 9 items (8 checklist items and one summary question). The mean total percent agreement (table 3.3) between the participants' and reference ratings for all items across the three papers and 9 checklist items was 92.9% (SD 0.391 between individual item ratings) and ranged from 83.3% (item 3- intervention tailoring) to 97.9% (items 2-intervention description and 5- schedule, duration, dose). The percentage agreement across all three papers was consistently lowest for item 3 (intervention tailoring), ranging from 81.25% for the first and second paper, and 87.5% for the third paper.

| CONSIDER Item | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| | % Agreement | % | % | % | % | % | % | % |
| 1. Pile, 2010 | 100 | 100 | 81.25 | 93.75 | 93.75 | 93.75 | 93.75 | 93.75 |
| 2. Tyler, 2011 | 87.5 | 100 | 81.25 | 87.5 | 93.75 | 87.5 | 87.5 | 87.5 |
| 3. Pollard, 2018 | 100 | 93.75 | 87.5 | 93.75 | 100 | 93.75 | 93.75 | 100 |
| **Mean % agreement** | 95.83 | 97.9 | 83.3 | 91.6 | 97.9 | 91.6 | 91.6 | 93.75 |
| SD | 7.22 | 3.61 | 3.61 | 3.61 | 3.61 | 3.61 | 3.61 | 6.25 |
| Mean total % agreement | 92.9 | mean variance between items (all) | 0.153 | Mean SD between items (all) | 0.391 | | | |
| **Cronbach's alpha** All items | 0.82 | | | | | | | |
| **Cohen's kappa** 16 raters/item | combined | 0.861 | Z = 91.9 | prob > z | **0.0000** | | | |

**Table 3.3:** Percentage agreement across CONSIDER items.

## CONSIDER Survey

Fifteen participants completed checklist survey (table 3.4). On average, participants reported being mostly comfortable with critical appraisal (score of 4, item 2) of trial publications, but only a little familiar with intervention fidelity (score 2 item 3), before participating in the reliability study. After reading the E & E materials and assessing the three papers, participants reporting being "moderately more (score of 3)" to "much more (score 4)" familiar with intervention fidelity (mean 3.4 item 4). On average, participants reported moderate difficulty appraising fidelity in the first paper they rated (average score 3, item 5), and "none" to "a little" difficulty by their third paper (average score 1.8, item 6). Across the 15 surveys, 13 of 15 participants reported they would be very likely to consider intervention fidelity when reading or appraising trial papers in future (average score 3.9, item 7). Fourteen of the fifteen participants reported finding the CONSIDER checklist a useful tool for evaluating intervention fidelity in complex intervention clinical trial papers, with one participant not commenting on the usefulness of the tool.

All the participants reported finding the E & E paper a helpful resource when using the CONSIDER tool. Nevertheless, six participants reported having difficulty appraising at least 1 CONSIDER checklist item, with 7 participants reporting having difficulty appraising item 3, or whether trialists identified allowable intervention tailoring and modification as part of the study protocol, and 2 participants reporting difficulty differentiating item 3 from item 7, or whether modifications to the intervention were made during the trial.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| **Survey items** | Current Role? | Comfort level With critical Appraisal? | Prior familiarity with Intervention fidelity? | After using CONSIDER how familiar with fidelity? | With 1st paper, how difficult was it to use CONSIDER? | By 3rd paper, how difficult was it to use CONSIDER? | How likely to consider fidelity when reading trial papers now? | Any items Difficult to appraise with CONSIDER (number)? |
| **Participant** | | | | | | | | |
| **1** | 4 | 4 | 2 | 4 | 3 | 2 | 4 | 0 |
| **2** | 4 | 4 | 2 | 3 | 3 | 1 | 4 | 7 |
| **3** | 4 | 4 | 2 | 4 | 4 | 2 | 4 | 3 |
| **4** | 1 | 3 | 2 | 4 | 3 | 2 | 4 | 0 |
| **5** | 4 | 5 | 5 | 4 | 2 | 1 | 4 | 3 |
| **6** | 2 | 4 | 2 | 3 | 3 | 2 | 4 | 0 |
| **7** | 4 | 5 | 5 | 3 | 3 | 2 | 4 | 0 |
| **8** | 4 | 2 | 1 | 3 | 2 | 3 | 3 | 3 |
| **9** | 2 | 3 | 3 | 4 | 3 | 2 | 3 | 0 |
| **10** | 4 | 4 | 2 | 3 | 3 | 1 | 4 | 3 |
| **11** | 1 | 3 | 1 | 3 | 3 | 2 | 4 | 0 |
| **13** | 1 | 3 | 1 | 3 | 3 | 1 | 4 | 7 |
| **14** | 1 | 3 | 1 | 3 | 4 | 2 | 4 | 0 |
| **15** | 1 | 3 | 2 | 4 | 3 | 2 | 4 | 0 |
| **14** | 2 | 4 | 2 | 3 | 3 | 2 | 4 | 0 |
| **Mean** | | 3.5 | 2.1 | 3.4 | 3 | 1.8 | 3.9 | n/a |
| **Mode** | | 4 | 2 | 3 | 3 | 2 | 4 | 0 |

**Scoring:** Questions 2 and 3: Score 1: none, 2: A little, 3: Moderately, 4: Mostly, 5: Very
Question 4: Score 1: Same, 2: A little, 3: Moderately, 4: Much more
Questions 5-6: Score 1: Not, 2: A little, 3: Moderately, 4: Very

**Table 3.4:** CONSIDER survey

Participants comments were also collated and grouped thematically (table 3.5). The participants had mostly positive feedback about the CONSIDER checklist (tables 3.4 and 3.5). Comments identified helpful aspects of the checklist, such as the elaboration, explanation and scoring guide, and the checklist construct (fidelity in three interacting components: Design, delivery, and receipt). Reflecting the results of the percentage agreement analyses, items 3 and 7, addressing prespecified, allowable tailoring and adaption of interventions (item 3), and unplanned modifications to interventions or deviations from the protocol (item 7) were described as more challenging to assess. One participant described difficulty distinguishing between these, while another described difficulty determining when tailoring was being described. Another comment suggested modifying the description of item 3 to further clarify what kinds of adaptions, and by whom, would be included under "tailoring." A learning curve effect was also described, with 4 participants reporting that they found it easier to use the checklist and identify the criteria in the sample of papers after assessing the first and second papers. This was expected, as the participants had not received any prior training with the checklist or in intervention fidelity as part of this study.

| Code | Feedback from participant |
|---|---|
| Overall feedback | "The detailed explanation and examples made it easy to follow and apply the items in regards to the different writing/journals' styles."<br>"The CONSIDER checklist is a useful tool that can easily assess clinical trials for fidelity."<br>"I was not really that aware of fidelity before this, but now wonder why I don't hear about it more. The checklist is very helpful but I think also raising questions about a study generally. We discussed this in our next journal club." |
| Helpful aspects of the checklist | "Construct was easy to follow."<br>"The attached overview, key terms and scoring examples and elaboration were helpful."<br>"The examples and explanations were very helpful."<br>"It helped to have definitions for fidelity and adherence. There were different words used to describe similar concepts in the papers. It helped that they also in the checklist at examples." |
| Items that were easy to assess | "The first 5 items in the scoring checklist are easy to follow, understand and to look for while reading the papers. Number 6, 7 and 8 require a more careful look at the data. "<br>"For the items 1, 2, 4, 5, and 6 are easy to appraise." |
| Items that were hard to assess | "I found tailoring somewhat difficult to determine."<br>"It was initially challenging to separate the concept of tailoring and modification. "<br>"Some parts of the articles are in grey area and hard to use this tool to evaluate. Especially for items 3 and 7, most of the article are not directly mention the involved modification."<br>"I found the articles not related to my field (Speech and Language) somewhat difficult to understand and therefore more difficult to appraise."<br>"Overall, I find scoring easier when more objective measures are present, leaving less to my personal subjective judgement." |
| Suggested modifications | "For number 3, the tailoring, it asks "what adaptions may be made?" This made me wonder if you meant adaptations made by the therapists, or the patients, or the people running the study? I think it would help if this could be clarified."<br>"If the research is not including home program exercise, the description of the adherence could be modified, such as the attendance of each session."<br>"Perhaps an entire article could also be used as an example prior to scoring a first article. This practice article would help ensure the scorer can find all the relevant items and score them appropriately." |
| Learning curve | "With the first article, the items were more difficult to appraise even with the detailed examples and explanations due to their non-sequential order in the articles. By the third article, it was easier to find the items or their absence and score accordingly."<br>"Every paper had differences in style, reported information, which made it necessary to take the concepts in the Checklist score and apply them in different ways. It got easier with practice."<br>"With practice, it became easier to use the checklist, and easier to see fidelity." |

**Table 3.5:** User comments and suggestions

## Revised checklist

As a result of the participant survey responses, minor modifications were made to the checklist (figure 3.4). Item 3 (tailoring) was modified to clarify that it refers to allowable adaptations that can be made by the interventionist (researcher, investigator, therapist, treatment provider) to the intervention when tailoring is needed for individual participants. Examples of reasons for tailoring were added to the description. Additional examples of tailoring from published trials were also added to the explanation, example, and scoring guide. The wording of item 6 was also modified to clarify participant adherence. One participant's survey response recommended creation of a training set of at least one trial paper and accompanying CONSIDER assessment. Such a training set can be developed for future users of the checklist, complementing the elaboration, explanation and scoring guide.

| Items to indicate whether fidelity supported? | Items 1-5 Scoring: 0: Absent 1: Somewhat discernible 2: Mostly discernible |
|---|---|
| **Item 1.**<br>CONSIDER Design<br>TIDieR item 3<br>CERT item 1 | **Materials**: Describes physical or informational materials or methods used to train intervention providers in study methods or intervention delivery, or to train or help participants in carrying out intervention activities. E.g.: Intervention manuals, videos or instructional aids, exercise sheets, etc. |
| **Item 2.**<br>CONSIDER Design, Delivery<br>TIDieR 4<br>CERT 8,9,13, SPIRIT 11a | **Detailed description of the intervention**: Describes what intervention procedures, processes, or activities providers carried out. Describes what the experimental intervention(s) looks like, includes, or how performed. |
| **Item 3.**<br>CONSIDER Design<br>TIDieR 9<br>CERT 14, SPIRIT 11b | **Tailoring:** If intervention was planned to be, or allowed to be, personalised, titrated or adapted for individual participants during the trial: describes why, when, or how adaptations may be made. For example, how or when might exercise intensity or duration be increased or intervention progression occur? Can an intervention be modified for persons with mobility impairments or mobility aids, or delivered remotely vs in person? |
| **Item 4.**<br>CONSIDER Design, Delivery<br>TIDieR 5,<br>CERT 2: Who | **Intervention providers:** Describes their qualifications, background, expertise, and any training given in the intervention or study procedures.<br>**Score: 1** if reports one of either the providers' qualifications/expertise OR the training provided for the study.<br>**2** if both described. |
| **Item 5.**<br>CONSIDER Delivery, Receipt<br>TIDierR 6, 8<br>CERT 3,4,7,8, 9,12, 13<br>SPIRIT 9, 11a | **Schedule, duration, intensity, or dose:** Reports the number of times the intervention was delivered, or meant to be, and over what period of time. Was intervention(s) performed individually or in a group, supervised or not, and where performed (e.g.: home, community, clinic...)? |
| **Was fidelity monitored?** | **Individual item scoring** |
| **Item 6.**<br>CONSIDER Delivery and Receipt<br><br>TIDieR 11<br>CERT 5<br>SPIRIT 11c | **Fidelity and adherence monitoring:** Describes strategies planned or used to monitor, maintain, or improve participant adherence (attendance in intervention sessions or undertaking tasks or interventions required for the study) or intervention delivery (or performed correctly) as intended in the study protocol or manual. May include terms such as: fidelity, integrity, adherence, compliance, per protocol.<br><br>May describe how these were monitored. Example: fidelity checklists, audit of session notes, video or audio recording of intervention sessions, supervision during intervention, participant logs, diaries, worksheets, etc.<br><br>**Score: 0** if no monitoring (as above) is reported.<br>**1** if only participant adherence monitored (e.g.: frequency of attendance ,exercise, logs, dropout rate).<br>**2** if only intervention fidelity (intervention delivered as per study protocol or procedures) is monitored.<br>**3** If both adherence and intervention fidelity (as defined above) are monitored. |
| **Was fidelity reported?** | **Individual item scoring** |
| **Item 7.**<br>CONSIDER Delivery<br><br>TIDieR 10<br>CERT | **Modifications:** Was the intervention modified during the course of the study, or were deviations from protocol reported, including changes to interventions, unintended participant cross-over between groups/interventions?<br>Score: **(-2)** Unplanned deviations from the protocol are reported, including changed treatment frequency, or cross over between intervention/control groups, OR less than 75% fidelity achieved (see item 8)<br>**0** if paper does not report/discuss modifications that took place during trial interventions, or not reported and cannot be calculated.<br>**1** if tailoring or modifications made, were consistent with the study protocol or intervention manual, or 75%-90% fidelity reported (or can be calculated) if assessed.<br>**2** if authors report no, or very few, unintended intervention modifications or deviations from protocol were made. |
| **Item 8.**<br>**Was fidelity maintained?**<br>CONSIDER Delivery, Receipt<br><br>TIDieR 12<br>CERT 5, 16 | **Fidelity assessment:** If the paper reports that participant adherence or intervention fidelity were assessed or monitored: how much, or to what extent, where these achieved? Paper reports:<br>**Score: (-1)** If participant adherence or intervention fidelity were measured and either poor (<75%) or not achieved, or (-2) was scored for Item 7.<br>**0** if no report of adherence measures or fidelity assessment or results are found in the paper (Impossible to determine).<br>**1** if only participants adherence reported, > 75%<br>**2** if intervention fidelity (interventions delivered as intended is achieved (>75%), but adherence is Not reported in the paper, cannot be calculated, or was not achieved (75% or more).<br>**3** if both adherence and intervention fidelity are reported to be achieved (75% or more). |

Figure 3.4: Revised CONSIDER checklist

## DISCUSSION

This study investigated the CONSIDER checklist's inter-rater agreement, internal consistency, and inter-rater reliability across 3 clinical trial papers and 552 item and survey ratings by 15 independent raters (participants). CONSIDER is a reliable, internally consistent checklist for assessing intervention fidelity in complex intervention clinical trial papers. Overall, percentages of agreement for the checklist items were high, with raters often choosing the same response options when assessing fidelity in the selected trial publications. The checklist items also have good reliability and internal consistency (the degree to which items on the checklist are correlated with one another), with Cronbach's $\alpha$ of 0.82. A high Kappa coefficient (0.861) confirmed the checklist's inter-rater reliability and the extent to which the items described in the checklist are likely correct representations of the variables being measured. The high kappa coefficient also indicates that individual items could be distinguished between papers when participants used the checklist.[400]

Although percentage of agreement was high 92.9% (range 81%-100%, mode 93.3%), users of the checklist achieved the lowest percentage agreement and reported difficulty assessing item 3-planned intervention tailoring to individual participants, and item 7, unplanned modifications made to interventions during the trial, or differentiating between them. Tailoring and adaptation of interventions during clinical trials are often poorly documented and reported in the body of literature for complex interventions[12,145,392,432], despite their inclusion in the Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) statement (item 11b: criteria for intervention modifications)[26], TIDIER checklist (item 9: tailoring and item 10: modifications) in

2014[150], and CERT (item 14: tailoring and 16: how well-planned vs actual) in 2016.[151] A number of systematic reviews of complex intervention implementation have found poor reporting of intervention tailoring or adaptation.[46,199,200,396,433]

A meta-analysis of reporting of experimental and control therapies in 215 Stroke rehabilitation trials by Lohse et al. (2018) found that reporting of how interventions were tailored to individuals or modified during trials was generally poor, despite the wide range of participant differences in functional ability, age, and cognitive status within trials.[199] Similarly, a systematic review of sports injury trials found only 12% of all included studies reported aspects of intervention tailoring or modification.[433] Ribeiro et al.'s (2021) appraisal of clinical trials included in a Cochrane review of exercise therapy for the management of shoulder pain also found that only 2 of 10 trials reported any (only partial) information about whether interventions were modified, and none reported pre-determined allowable intervention tailoring or adaptation.[396]

These findings and the participants' experience with items 3 and 7 may also reflect the inherent challenges of defining[332], documenting, and reproducing complex interventions.[5,7] Interventions in physiotherapy, speech-language therapy and other rehabilitation domains are highly individualized in their application and delivery.[5,7] It can be difficult to predict or report precisely how complex interventions may be, or have been, tailored for individual participants or settings.[12] Pre-specifying allowable tailoring and adaptation also requires understanding of interventions' essential "active ingredients," or the treatment elements producing patient outcomes, and differentiating them from intervention components that can be modified without

substantially altering the intervention's mechanisms of action or patient outcomes.[204,262,392]

This can prove challenging in complex interventions such as those in rehabilitation, which have frequently been compared to a "black box" due to largely poorly-defined essential characteristics such as treatment theory and intervention active ingredients and their effects on intervention outcomes.[330,332] Empirically validated taxonomies for rehabilitation interventions have been developed to help researchers and practitioners document and identify interventions' active ingredients.[330,332] Their uptake may facilitate documentation and assessment of intervention tailoring and adaptation in future.[330] Nevertheless, it is essential to assess not only what was delivered during a trial, but also whether (and why) it was modified, or whether unplanned elements were introduced to improve understanding of the factors influencing treatment outcomes and support the translation of evidence from clinical trials into clinical practice.[392] Items 3 and 7 in the CONSIDER checklist draw further attention to important aspects on in-trial intervention delivery that may greatly influence treatment outcomes, research replication and implementation, and continue to require greater attention.[204,392]

## Strengths, limitations, and future directions

The reliability study has several strengths. Following the Guidelines for Reporting Reliability and Agreement Studies (GRRAS)[400] recommendations, we calculated a sample size needed for this study and recruited a sample of participants who are representative of future users of the checklist. The checklist was developed through extensive systematic review and thematic synthesis of data from a large dataset of methodological and clinical trial papers in complex

interventions (chapter II), following guidance for developing and evaluating fidelity checklists.[132] We also developed an extensive explanation, elaboration and scoring guide with illustrative examples of checklist items and their scores from our previously compiled database of complex intervention clinical trial papers. Participants also assessed three randomly selected papers that are representative of papers the checklist would be used with in future.

## Limitations

It was not feasible to train the participants to use the checklist before they assessed the papers. Although we did not train participants to use the checklist or intervene in their assessments, the internal consistency, reliability, and inter-rater reliability of the checklist and participants were high. However, it is possible that some experience in using the checklist before assessing the sample of studies would have improved participants' ease in using it, particularly with items 3 and 7. It is also possible that participants' years of research or critical appraisal experience may have influenced their skill in using the tool. However, we attempted to overcome this by recruiting participants with a minimum amount of either clinical or research experience.

Very high Cronbach's $\alpha$ values (>90%) indicate internal consistency but may also reflect item redundancy in checklist with a number of items.[434] As a result, it may be that some checklist items are unnecessary or redundant when $\alpha$ is very high.[434] Item redundancy was built-in to the checklist, by design. A wide range of terms and constructs for fidelity and its components are found, and often used interchangeably, in complex intervention literature. Considering these

factors, checklist items were required which could capture intervention fidelity in actions reported in trial papers in a variety of ways.

The CONSIDER checklist was also designed to be a multicomponent assessment of fidelity across three inter-related components of fidelity: intervention design, delivery, and receipt. Some overlap between items was expected, reflecting the complex nature of intervention fidelity. Intervention fidelity is increasingly viewed as a multi-faceted concept encompassing a range of interacting participant and practitioner actions and behaviours.[12] For example, the quality of intervention delivery (fidelity), may be influenced by the provider's level of experience or training in the intervention before the trial begins, monitoring to prevent intervention drift during the trial, as well as the degree to which key intervention details have been operationalized, among other factors.

There is also some overlap between items assessing reporting and items assessing methodological quality or processes (i.e. fidelity) in the CONSIDER checklist. For example, while item 6 (Fidelity monitoring: Describes strategies planned or used to monitor, maintain, or improve participant adherence or intervention delivery as intended in the study protocol or manual) includes an emphasis on reporting quality, or whether the paper reports the strategies used to monitor fidelity, items 7 (Modifications: Was the intervention modified during the study, or were deviations from protocol reported, including changes to interventions, unintended participant cross-over between groups/interventions?) and item 8 (Fidelity assessment: If the paper reports that participant adherence or intervention fidelity were assessed or monitored:

how much, or to what extent, where these achieved?) include a mix of whether fidelity was maintained and whether it was reported. This redundancy was purposeful and intended to maximize the checklist's ability to capture information about fidelity in trial publications, given the wide range of terms and descriptions used for fidelity in complex interventions literature.

However, this strategy also has its limitations, reflecting the difficulty of assessing methodological quality in non-pharmaceutical RCTs.[435] It may not be possible to determine whether checklist items (representing study processes or procedures) assessed by the checklist were conducted, but not reported, or not conducted and not reported. Well-conducted trials may be reported badly, and studies with methodological flaws may be reported inaccurately, obscuring deficits in methodologic quality. This is an acknowledged limitation of tools assessing study quality or methodological characteristics in trial publications, including the CONSIDER checklist.[70,435] This issue is explored more extensively in Chapters IV, V, and VI of this thesis.

The overlap between study conduct and study reporting creates limitations in this current study, including how methodical characteristics not reported should be scored. A scoring system was needed for the checklist to be able to achieve its purpose in this thesis. An extensive review of literature was conducted to inform the development of the checklist, which included careful consideration of other leading checklists assessing fidelity in trial publications, such as Gearing's CIFG (psychology and behavioural health). The CIFG assesses intervention fidelity across 22 items that are scored as 0-absent/minimal, 1-moderate, 2-extensive.[40] As a result, the decision was made to score items that were reported with a positive or negative score, depending on the

degree to which the items were achieved, while items that were not reported (could not be identified in the trial publication) were assigned a score of "0," rather than a non-numerical score (e.g. not reported, not applicable). It is possible that a "0" score, rather than another option (e.g. not-applicable, not-reported, or a scale from 1-5, rather than "-2" to "+2" and including "0") may bias the results of assessments with the checklist in papers with poor fidelity reporting. Future development of this checklist through Delphi Consensus can further explore these issues and the checklist scoring system.

High internal consistency does not necessarily reflect a checklists' validity, however.[434] A checklist may be reliable but not valid. It was not the purpose of this study to establish the validity of the CONSIDER checklist. However, the checklist was developed through extensive systematic review, Best-Fit Framework Synthesis, and thematic synthesis of a large dataset of methodological and clinical trial papers in complex interventions (chapter II). It was developed specifically for use in the context of providing a transparent, reproducible way to assess and document intervention fidelity in complex intervention randomized controlled trials for the meta-epidemiological study in the following chapter (Chapter V), rather than for broad adoption by a range of users. As such, it does not represent the results of consultation with a wide group of expert trialists or clinical trial methodologists. That was beyond the scope of the checklist's initial development.

## Future directions

Intervention fidelity has been identified as a component of performance bias and an important

criterion for assessing risk of bias in interventional studies. It is rarely described or included in risk of bias or study quality assessment scales, however. Findings from complex interventions trials, or from systematic reviews and meta-analyses of them, that were administered with poor fidelity may be unreliable for informing clinical decision making and may hinder the translation of effective treatments from clinical trials to clinical practice.[290,436,437] The CONSIDER checklist could be used, or developed further, to complement risk of bias assessments such as the Cochrane Risk of Bias (ROB-2) by providing more targeted assessment of intervention fidelity in systematic reviews and meta-analyses.

Future development and validation of the checklist may also increase its utility for grant reviewers, funding bodies, and journal editors evaluating treatment outcome studies. The National Institute for Health Research (NIHR), for example, requires researchers to create a monograph on completion of funded studies, including detailed description of intervention fidelity in the research.[177] The CONSIDER checklist can facilitate evaluation of intervention fidelity in study monographs such as those required by the NIHR.

## CONCLUSION

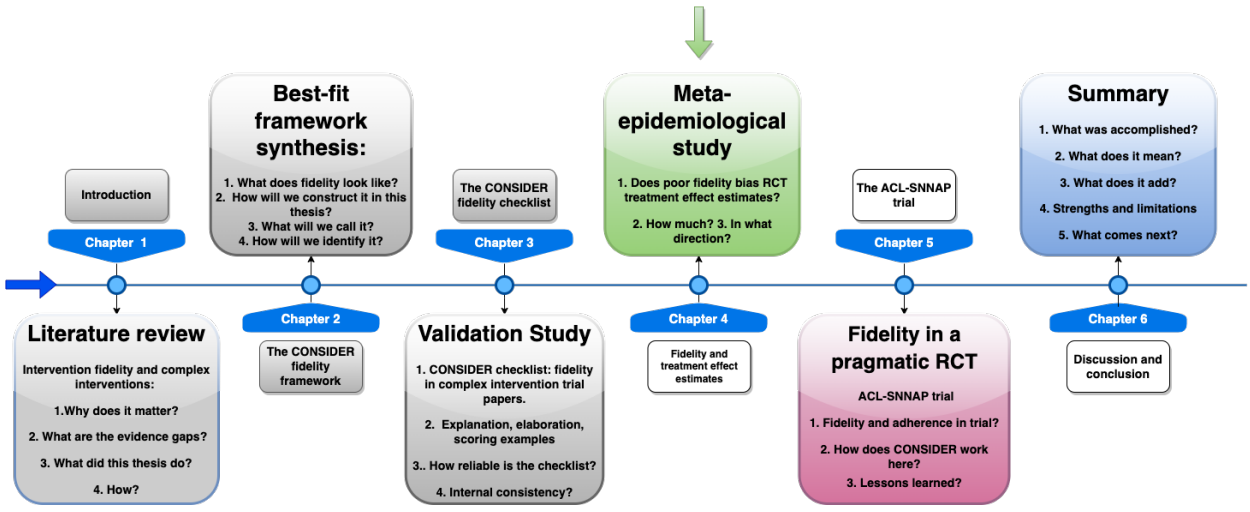
CONSIDER is, to the best of our knowledge, the first checklist developed to assess fidelity in trial publications for physical complex interventions and one of only a handful of checklists developed for assessing fidelity in published papers. It was developed specifically for use in this DPhil thesis to provide a transparent, reproducible way to assess intervention fidelity in complex intervention randomized controlled trials for the meta-epidemiological study in the following chapter

(Chapter IV). However, CONSIDER can nevertheless make an important contribution to the interpretation of findings from complex interventions trials and systematic reviews and meta-analyses. Future development and validation of the checklist may also increase its utility for grant reviewers, funding bodies, and journal editors evaluating complex intervention treatment outcome studies.

Chapter IV: The influence of intervention fidelity on treatment effect estimates in clinical trials of complex interventions: A meta-epidemiological study.

## Chapter Summary:

In this chapter, systematic searches are undertaken to identify meta-analyses of randomised controlled trials (RCTs) in physiotherapy, speech-language therapy, exercise, and physical activity interventions. A meta-epidemiological study is undertaken to determine whether intervention fidelity influences the magnitude and/or direction of treatment effect estimates derived from randomized controlled trials of physical complex interventions. The primary aims of this chapter were to determine if poor intervention fidelity biases treatment effect estimates derived from complex intervention RCTs, identify the direction of that bias (toward or away from the null hypothesis), and quantify the bias in the sample of meta-analyses.



## BACKGROUND

Randomized clinical trials provide the most reliable method for assessing the effectiveness of therapeutic interventions.[438–440] However, flaws in their design and conduct may bias trials' treatment effect estimates, leading to over or underestimation of the true intervention effect.[57,82,439] Characteristics of clinical trials, such as blinding or intervention fidelity, may create systematic differences in intervention effect estimates between trials with and without those characteristics.[57,439] Poor fidelity may lead to variations in intervention delivery resulting in differences in participants' exposure to intervention components (dose).[63] This may lead to variability in outcome achievement, inflating error variance in post-test outcomes and decreasing trials' statistical power to detect significant effects.[53,63,89,90] Pooled data that includes biased treatment effect estimates can also bias the results of meta-analysis used to support evidence-based clinical decision making, leading to ineffective interventions being implemented into practice, or effective interventions not being implemented.[57,81,439,441,442]

Meta-epidemiological studies are designed to investigate the impact of study-level characteristics, like fidelity, by investigating the association between the and the intervention effect estimates from collections of meta-analyses.[78] By contrasting the results of trials with a characteristic of interest with the results of trials with without that characteristic, they provide estimates of the average bias associated with that characteristic.[79,80] Despite the rapid growth in meta-epidemiological research in the past two decades, considerable methodological variation subjects meta-epidemiological studies to increased risk of bias and has made interpretation of their results challenging.[57,80,443] Imprecise or ambiguous definition of trial characteristics and

heterogenous trial samples are frequently seen in meta-epidemiological papers.[80,444] Other limitations include infrequent use of power and sample size calculations for meta-epidemiological studies,[80,436] resulting in insufficient sample sizes and poor power,[445] and lack of exploration of variation in the impact of trial characteristics on effect estimates between meta-analyses or between trials within meta-analyses.[80,446] Additionally, few meta-epidemiological studies examine continuous outcomes, despite their frequent use in clinical research.[82] This may limit the application of findings from meta-epidemiological studies of dichotomous outcomes to trials with continuous outcomes. These factors may influence the variations seen in the results of different meta-epidemiological studies of the same trial characteristics.[80]

Since the first published meta-epidemiological study in 1995,[447] 58 trial characteristics have been examined in 166 papers including meta-epidemiological analysis.[80,448] To the best of our knowledge, none have investigated the influence of intervention fidelity on treatment effect estimates in RCTs. Two meta-epidemiological studies have included a component of fidelity as a criterion of interest.[441,449] Van Tulder et al. examined associations between internal validity and treatment effect differences in 216 RCTs of low-back pain interventions. "Compliance" was included as one of 11 characteristics of interest investigated. Investigators determined if participants' compliance to trials' interventions were acceptable based on the reported intensity, duration, and frequency of sessions for experimental and control interventions.[441] No significant difference in effect sizes was found between trials with low or high levels of compliance (difference of -0.01, 95% CI: -0.15 to 0.14 when compliance was high). However, no parameters

or definition for compliance are provided in the published paper, making it impossible to ascertain how compliance was conceived or differentiated from other aspects of participant performance.

Hempel et al. investigated associations between trial quality and estimates of treatment effect in in four datasets, including Van Tulder's [441] sample, totalling 481 RCTs across a range of clinical fields.[449] Eleven quality criteria were investigated, including "similar cointerventions," and "acceptable compliance." No statistically significant difference in effects size was found between trials with high or low levels of compliance when adjusted for size of the treatment effect, the condition being treated, the type of outcome, and variance in effect sizes.[449] As in the study by van Tulder however, adequacy of compliance was a characteristic of interest but compliance was not defined, nor were criteria for assessing it. A clinician with trial research experience was contacted to establish reasonable compliance rates in included RCTs, but how this was determined was not described.

Compliance is also a multi-faceted term that may encompass a variety of concepts related to fidelity, including the extent to which: participants follow the instructions they have been given for participating in a clinical trial, adhere to prescribed behaviours or treatment dosages, avoid proscribed behaviours or activities, attend the prescribed frequency of intervention sessions or achieve the prescribed participation level for experimental or control interventions.[2,16,40,441] A number of other descriptions for compliance were identified in chapter II. In both Van Tulder and Hempel's studies, lack of specification of compliance and which behaviours it represents

limits interpretation of their findings on the influence of fidelity on treatment effect estimates in RCTs.

Poor intervention fidelity and protocol violations increase variability with which interventions are administered in clinical trials. This variability is likely to artificially increase or diminish the results that are obtained, leading to Type I or II errors.[23] To the best of our knowledge however, the influence of intervention fidelity on treatment effect estimates in clinical trials remains unquantified. This meta-epidemiological study examined meta-analyses of randomized controlled trials in three disciplines representative of complex interventions in healthcare: physiotherapy, speech-language therapy, exercise, and physical activity interventions. These disciplines yielded the greatest number of eligible papers in the systematic review of fidelity constructs and best-fit framework synthesis completed in earlier stages of this thesis (Chapter III). Intervention fidelity's influence on treatment effect estimates derived from complex intervention RCTs can significantly influence the interpretation and appropriateness to change practice of clinical studies of complex interventions and clinical practice guidelines based on systematic reviews of clinical trials.

## Aims and Objectives

**AIM I:** To quantify the prevalence of fidelity monitoring and intervention fidelity in RCTS of complex interventions in the physical domain (rehabilitation, broadly defined).

**AIM II:** Determine whether intervention fidelity influences (biases) the estimates of treatment effects derived from RCTs of physical complex interventions, and in which direction.

Hypotheses:

a. Poor intervention fidelity quantifiably biases treatment effect size and precision in RCTs of physical complex interventions.

b. High degrees of intervention fidelity are associated with a larger treatment effect estimates and greater precision, while poor fidelity supresses treatment effects and reduces precision in complex intervention RCTs.

AIM III: To determine whether reporting of intervention fidelity depends on a trial's outcomes.

Hypothesis: Intervention fidelity is reported more often in trials with outcomes favouring the experimental intervention.

Objectives:

1. To systematically search for meta-analyses of randomized clinical trials in three physical complex interventions: physiotherapy, speech therapy, and exercise-physical activity interventions. (AIM I, II)
2. To quantify and characterize intervention fidelity monitoring and reporting in RCTs in eligible meta-analyses. (AIM I)
3. To investigate how sample size and risk of bias influence the relationship between intervention fidelity and treatment effect estimates through subgroup analyses.
4. To evaluate the influence of intervention fidelity monitoring on estimates of treatment effects in meta-analysed RCTs in a paired-design, meta-epidemiological (meta-meta-analytic) study. And in which direction? (AIM II)

## METHODS

A systematic search for meta-analyses (AIM I) in complex interventions and a meta-epidemiological study (AIM II) were conducted to investigate the association between intervention fidelity and treatment effect estimates in meta-analyses of complex intervention RCTs.[78]

The meta-epidemiological study was developed following the recommendations of Sterne et al. in "Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological research[79]" and Moustgaard et al., whose "Ten questions to consider when interpreting results of a meta-epidemiological study—the MetaBLIND study," provides guidance for the critical interpretation of meta-epidemiological studies. The study is reported along the guidelines for meta-epidemiological research proposed by Murad and Wang.[450]
The protocol was initially submitted for registration on Prospero, then registered on the Open Science Foundation protocol registry (https://osf.io/z83wc).

**Objectives I-II:** A systematic search for meta-analyses of RCTs in physiotherapy, speech therapy, and physical activity promotion or exercise interventions, with at least two treatment arms and reporting quantitative data for continuous outcomes, published between 2010 and 2020, was undertaken following the recommendations of the Cochrane Handbook for Systematic Reviews of Interventions.[70]

In order to determine the minimum number of meta-analyses required to achieve statistical

power for detecting the hypothesized mean impact of fidelity on treatment effect estimates, a power calculation was undertaken with the Sorbonne-Paris Centre of Research in Epidemiology and Statistics Power and Sample Size calculator for meta-epidemiological studies (http://www.clinicalepidemio.fr/metaepidemio/).[451] Using an estimated mean impact of fidelity on treatment effect estimates β= 0.30[452], assumed variability of the impact between trials within meta-analyses ($\kappa^2$ = 0.03) and between meta-analyses ($\phi^2$ = 0.002) following the guidance of Giraudeau, et al.[436] a mean number of trials with the characteristic of interest (4) and without the characteristic of interest (6) based on a systematic review of fidelity in clinical trials of complex interventions in previous stages of this program of research (CONSIDER), a minimum of 14 meta-analyses are required to have 80% power to detect a mean difference β = -0.25 with α = 0.05. To accommodate the possibility of greater variability or lower mean difference between trials than was anticipated, 19 eligible meta-analyses, and at least 190 RCTs were sought.

## Search Methods

Electronic searches were conducted without language restrictions in PubMed with keywords and MeSH terms for physiotherapy, physical therapy, rehabilitation, exercise, physical activity, expressive language, receptive language, speech therapy, language therapy, articulation, interventions, continuous outcome, meta-analysis, systematic review, and others (appendix IV, figures 4.1-4.4). The search strategy combined validated filters related to meta-analyses.[453] Potentially eligible meta-analyses were screened by title and abstract and then by full-text based the following criteria:

Inclusion criteria: Meta-analysis level

1. Must be Meta-Analyses (MA), identifying and summarizing evidence from published reports through quantitative analyses.
2. Must analyse RCTs in physiotherapy, speech-language therapy, and exercise-physical activity promotion interventions.
3. Examine at least one continuous outcome.
4. Must include a minimum of 4 RCTs and provide quantitative data for treatment effects.
5. RCTs included in the selected MAs must:
    a. evaluate a therapeutic intervention.
    b. have a placebo, wait-list, no intervention, or other intervention control.
    c. be reported in sufficient detail for effect sizes to be calculated.

Exclusion criteria: Meta-analysis level

1. Systematic reviews without meta-analysis.
2. Meta-analyses of economic impact studies or interventions outside of physiotherapy, speech-language, physical activity, or exercise interventions.
3. Meta-analyses of less than 4 RCTs
4. Observational, non-randomized, or uncontrolled trials.
5. Trials reporting only dichotomous outcomes.

Inclusion criteria: Meta-epidemiological analysis level

1. Meta-analyses: Assessment of all RCTs in a meta-analysis must result in at least 2 of the 2 RCTs meeting criteria for fidelity as follows:
    a. Minimum scores of 1 on CONSIDER checklist item 6 (Was fidelity monitored), or

b. Minimum score of 1 on CONSIDER checklist items 7 or 8 (was fidelity achieved), or
c. Minimum score of 6 on "was fidelity supported?"

The Complex Interventions Design, Delivery, Receipt (CONSIDER) intervention fidelity checklist was developed in Chapter II of this thesis (figure 4.1). The checklist was piloted, and its reliability was established in the reliability study in Chapter III.

Exclusion criteria: Meta-epidemiological analysis level

1. Meta-analysis does not include RCTs meeting inclusion fidelity criteria with the CONSIDER checklist.

Titles and abstracts yielded by the search were screened the against the review's inclusion criteria. Papers meeting eligibility criteria by title and abstract and papers whose eligibility could not be determined by title and abstract were reviewed by full text. Trial protocols and registrations were also reviewed, when available, for additionally information about eligibility or intervention fidelity. When necessary, study authors were contacted for additional information to resolve questions about studies' eligibility or trial conduct.

## Data extraction

Meta-analysis level*:* Once eligible meta-analyses were identified after title-abstract and full-text screening, they were exported into a CSV dataset with the following identifying variables: authors, year, title, article identifier (Pubmed), journal title. The CSV data set was then imported into the STATA 16.1 Statistical Package.[454] Random blocks of 25 meta-analyses were then

generated in STATA with the randomizr package[455]. Their constituent RCTs were extracted and assessed for fidelity with the CONSIDER intervention fidelity checklist by the thesis author (AP) to determine eligibility for the meta-epidemiological analysis (AIM II, Objective 3). If an insufficient number of meta-analyses meeting the fidelity criteria were identified in the first set of 25, the next sample of 25 meta-analyses was generated, and so forth, until 19 meta-analyses meeting the fidelity criteria were identified.

The following data was extracted from meta-analysis eligible for the meta-epidemiological analyses:

1. Bibliographic data (authors, title, and year of publication).
2. Healthcare discipline (Physio, Speech Therapy, Exercise or Physical Activity interventions).
3. Number of included RCTs.
4. Number of RCTs meeting fidelity inclusion criteria.

RCT level: Component RCTs in eligible meta-analyses were reviewed by full text and the following data was extracted:

5. Bibliographic data (authors, title, and year of publication)
6. Healthcare discipline (Physio, Speech Therapy, Exercise or Physical Activity interventions)
7. Sample size
8. Treatment effect estimate, or the mean change from baseline, and measures of variability (SDs and 95% CI:s) for the primary continuous outcome(s)

The primary outcome reported for each meta-analysis was used as the primary outcome for analysis. In MAs reporting two primary outcomes, data for both were extracted. If MAs reported multiple outcomes, the primary outcome was determined as the one with the largest number of trials.

Interventions within trials were classified as either experimental or control based on descriptions in the trial publications. Control interventions were classified as "no intervention," including wait-list control groups, "placebo," or "active control, including standard of care, treatment as usual, patient education or information, sham, or other therapeutic intervention.

## Fidelity and intervention reporting

Fidelity scores and items assessing intervention reporting (items 1-5) and fidelity (items 6-8) from the CONSIDER checklist were extracted for each RCT in included meta-analyses. A fidelity processes score was calculated for each RCT by adding scores for items 1-6 on the CONSIDER checklist (figure 4.2). These CONSIDER scores were also re-coded as categorical variables for quantitative analyses with the statistical processing software (STATA 16.1) in the meta-epidemiological analyses, as:

0- fidelity monitoring absent
1- only participant adherence monitored
2- only intervention fidelity monitored
3- both participant adherence and intervention fidelity monitored

## Risk of bias

Risk of bias was evaluated with the Cochrane Risk of Bias Tool (RoB).[439] Studies were classified as low risk, some concerns (unclear), or high risk of bias.

## Data Analyses

All data analyses were prespecified in the study protocol. Descriptive statistics were used to quantify the number of eligible meta-analyses, eligible RCTs, total number of study participants across trials, and the number of RCTs reporting participant adherence, intervention fidelity monitoring or intervention fidelity (Objectives I-II). The prevalence of fidelity and fidelity monitoring across eligible RCTs was calculated as:

$$\frac{\text{total number of RCTs meeting fidelity criteria}}{\text{total number of RCTs in the meta-epi sample}} = \text{Fidelity prevalence}$$

Further analyses was undertaken to investigate whether reporting of fidelity depended on a trial's outcome.[456] For example, researchers in trials resulting in small or non-significant treatment effect estimates may report on intervention fidelity to help explain the trials results, while researchers in trials resulting in large treatment effect estimates may not report the intervention fidelity monitoring in the study. This was investigated through two analyses. In the first, all RCTs were divided into two groups. The first group included those with either a small or null treatment effect, or standard mean differences (SMD) of < 0.50 as categorized by Cohen.[457] The second group included those with a moderate to large treatment effect ≥ 0.50. In the second analysis, all RCTs were sorted by whether the RCT found a statistically significant effect in

favour of the experimental intervention or for the control condition (including finding no significant difference between the two). The number of RCTs reporting intervention fidelity monitoring, and the number reporting maintaining fidelity was then compared between both groups in both analyses.

The **meta-epidemiological analysis** followed a two-level, paired design, in which treatment effect estimates were compared between trials with and without fidelity monitoring within each meta-analysis, and the mean impact of fidelity was then estimated across all meta-analyses (figure 4.1).[79,436]

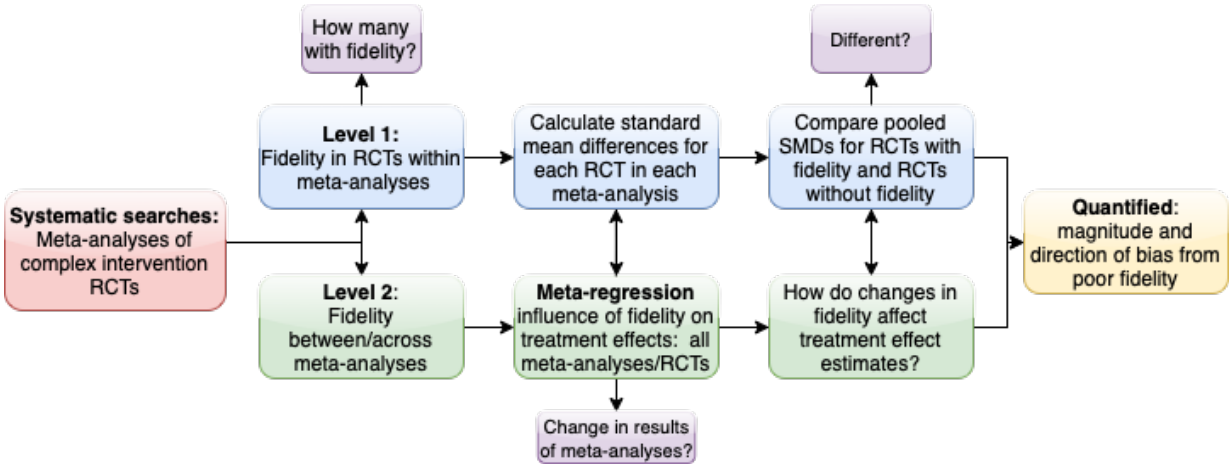



**Figure 4.1:** Meta-epidemiological study

First, individual RCTs within meta-analyses were sorted into those with or without intervention fidelity, and a two-level analysis was conducted using a meta-meta-analytic approach with an inverse-variance, Der Simonian and Laird random-effects model.[79,458] If an RCT appeared in more than one meta-analysis, it was analysed only in the meta-analysis with the fewest number of trials.

For the first level (within meta-analysis) analysis, standardized effect size estimates for the primary outcome of each trial were obtained.[459] Differences in mean outcomes between the experimental and control groups for the primary outcomes of each RCT were calculated and Standardized Mean Differences (SMD) calculated in RevMan 5.4 to accommodate for measurement of continuous outcomes with different scales between trials.[70]

With these SMDs, two pooled effect sizes were calculated for each meta-analysis. The first one corresponding to pooled effect size from studies meeting fidelity criteria and the other for those studies that did not. To maintain consistency across meta-analyses, endpoints were recoded if necessary, so that positive effect sizes indicted a beneficial effect of the intervention.[79] Comparison of trials with and without the characteristic of interest (fidelity) within meta-analyses maintains the grouping of trials created by authors of the meta-analyses, ensuring that trials being compared have broadly similar types of patients, interventions, and outcomes.[80]

## Statistical heterogeneity

Statistical heterogeneity within MAs was tested using the $t^2$, $Chi^2$ test (significance level: 0.1) and $I^2$ statistic, as follows[70,460]:

1. 0% to 30%: might not be important.
2. 30% to 50%: may represent moderate heterogeneity.
3. 50% to 90%: may represent substantial heterogeneity.
4. 75% to 100%: considerable heterogeneity.

### Effect of fidelity across all meta-analyses

To investigate the effect of fidelity across all meta-analyses, differences between pooled SMDs (dSMD) from trials with fidelity and trials without it for each meta-analysis were derived as:

$$dSMD = SMD_{no\ fidelity} - SMD_{fidelity}$$

The dSMDs were then pooled using Inverse-variance weighted meta-analyses with a Der Simonian and Laird random effects model to account for between-meta analysis heterogeneity.[458,461] As heterogeneity was expected within and between meta-analyses, the random effects model was used.[70,79,460] A dSMD <0 corresponding to more beneficial effects of interventions in trials with fidelity than trials without it.[79] The variability in bias estimates between meta-analyses was measured using $\tau^2$ as a measure of heterogeneity.

### Subgroup analyses

#### Risk of bias

Both high and unclear risk of bias judgements with the ROB have been associated with exaggerated treatment effect sizes in randomized trials.[462] Trials were sub-grouped by risk of bias on the ROB, and pooled meta-analyses performed for RCTs with low risk of bias, and for high and unclear risk of bias combined.

#### Sample size

Small study sample sizes (<100 participants) have also been associated with larger treatment

effect sizes, compared to larger sample sizes, in previous meta-epidemiological studies of RCTs.[82,463] It has been suggested that trialists' expertise and skill in recruiting participants may influence the trial's sample size, and may also influence characteristics such as participant fidelity to the intervention assigned, or loss to follow-up, influencing observed treatment effects.[443] In a second subgroup analysis, trials with less than 100 participants were characterized as small samples and those with greater than 100 participants as large sample sizes. Pooled meta-analysis of trials with and without intervention fidelity and these traits was then undertaken. Subgroup analysis was then performed for both risk of bias and study sample size.

## Meta-epidemiological analysis

In the **second-level analysis**, random-effects (Der Simonian and Laird) meta-regression was undertaken to test for interactions and explore associations between fidelity, sample size, risk of bias on the ROB, and treatment effect sizes.[70,436]

Meta-regression was conducted using the 'metareg' command in Stata 16.1 and weighted using effect size standard errors to measure the association between intervention fidelity, risk of bias, and study sample size and estimates of treatment effects. The dependent variable was the SMD for trials' primary outcomes, while the independent variables were intervention fidelity as a categorical variable (adherence only, fidelity, or fidelity and adherence), risk of bias as either a categorical (low, unclear, high) or dichotomous (low or unclear and high), and sample size as a categorical variable in 100 participant units.

### Impact of fidelity on the results of meta-analyses

Finally, the derived estimates of bias (meta-regression coefficients for fidelity and adherence present and categorical fidelity meta-regression coefficients) were used to estimate the effect of absent intervention fidelity on the results of individual meta-analyses. Within meta-analyses, the meta-regression coefficient for fidelity and adherence was subtracted from the SMDs of individual RCTs with absent fidelity and adherence (as evaluated with the CONSIDER checklist), and the meta-regression coefficient for fidelity as a categorical variable was subtracted for RCTs with adherence but absent fidelity (as evaluated with the CONSIDER checklist). An adjusted, pooled SMD was calculated for each meta-analysis. The resulting pooled SMD was then compared to the meta-analyses' published pooled effect to assess the impact that absence of fidelity may have had on the meta-analyses.

## RESULTS

**Objective I:** To systematically search for meta-analyses of randomized clinical trials in three physical complex interventions:  physiotherapy, speech therapy, and exercise-physical activity interventions.

### Characteristics of included studies

The search yielded 527 meta-analyses of physiotherapy, speech therapy, physical activity, and exercise interventions (figure 4.2). Of these, 125 and their constituent RCTS were assessed in 5, twenty-five meta-analyses groups until 19 meta-analyses were identified meeting fidelity eligibility criteria.

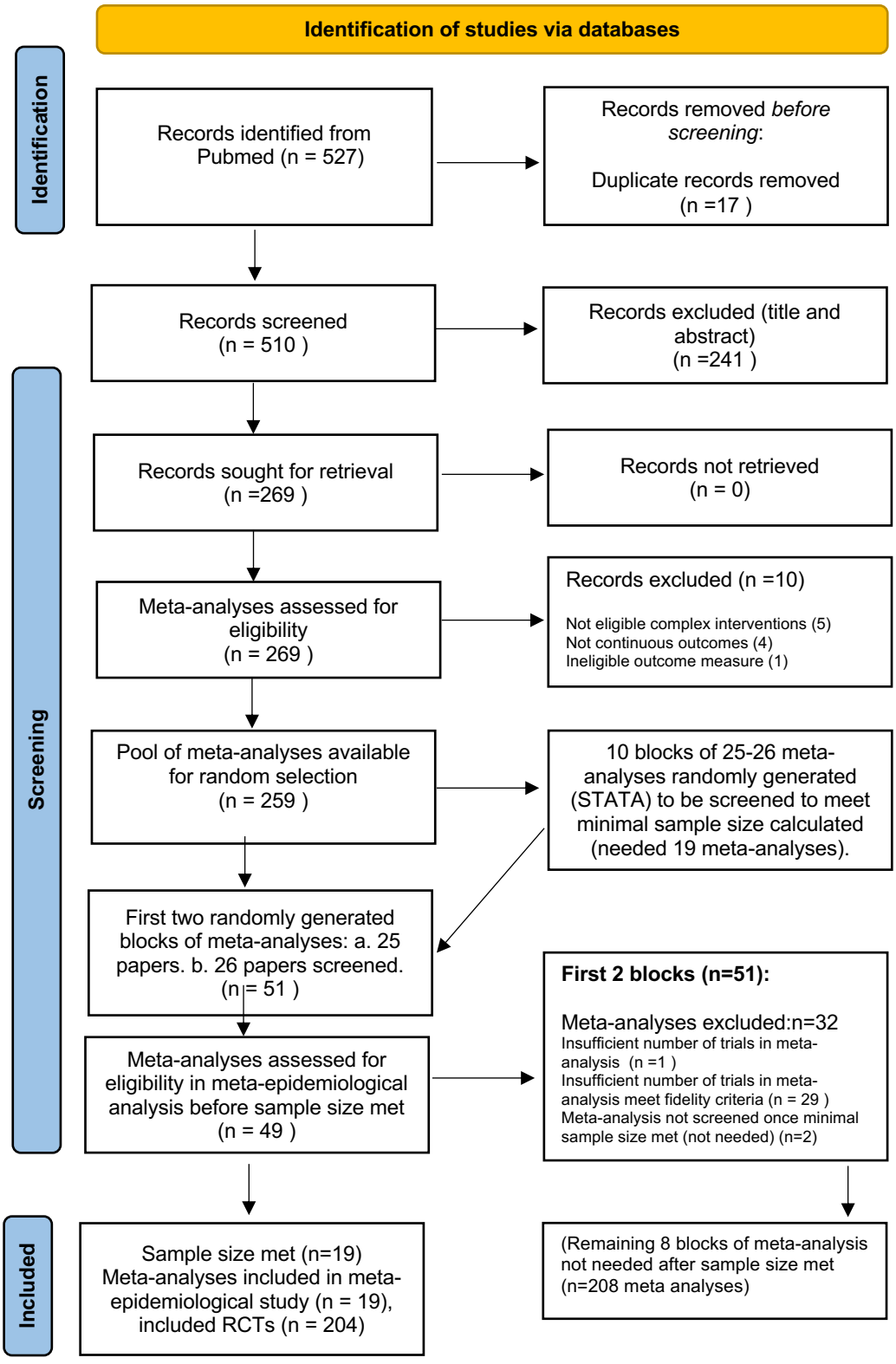



Figure 4.2: PRISMA Flow Chart

## Creating a sample of meta-analyses for the meta-epidemiological study

Ten blocks of 25-28 meta-analyses were randomly created by the STATA-randmomizr package. The first two blocks (51 meta-analyses) were screened, with 25 meta-analyses in the first and 26 meta-analyses in the second. Thirty meta-analyses were screened and 365 constituent RCTs assessed with the CONSIDER checklist until 19 meta-analyses meeting criteria for the meta-epidemiological analyses were identified. These included 204 RCTs, 248 randomized comparisons, and 25,056 participants (table 4.1). The median number of trials per meta-analysis was 10 (range 5-20), with a mean of 11 RCTs per meta-analysis. The median number of participants per meta-analysis was 806 (range 133-4442), with a mean of 118 participants per RCT. Eight meta-analyses examined physical activity or exercise interventions, 7 analysed interventions in Physiotherapy, and 4 analysed interventions in Speech Therapy.

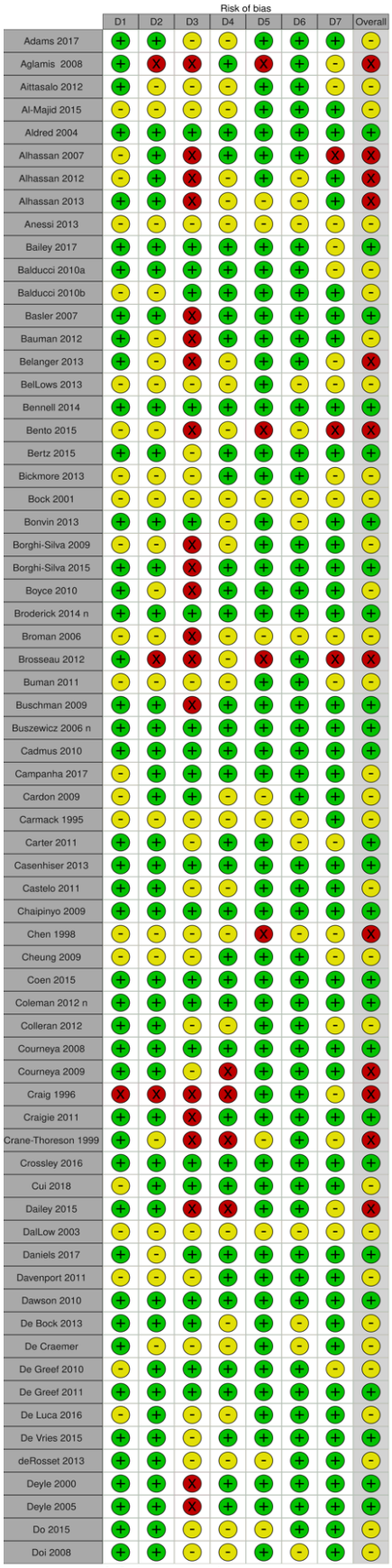
| Meta-Analysis | Year | Discipline | N RCTs | N with Fidelity | N Comparisons | Total Sample Size | Mean SS/RCT |
|---|---|---|---|---|---|---|---|
| Avery[464] | 2012 | Phys Act/Exercise Int. | 10 | 4 | 14 | 355 | 38.9 |
| Bellicha[465] | 2018 | Phys Act/Exercise Int. | 10 | 2 | 14 | 339 | 39.5 |
| Briani[466] | 2018 | Phys Act/Exercise Int. | 16 | 6 | 15 | 2693 | 171.5 |
| Cleave[180] | 2015 | Speech-Lang. Therapy | 7 | 4 | 8 | 153 | 19.1 |
| Depiazzi[189] | 2019 | Physiotherapy | 8 | 5 | 9 | 415 | 60.1 |
| Finch[467] | 2016 | Phys Act/Exercise Int. | 16 | 6 | 16 | 4422 | 276.4 |
| Hampton[398] | 2016 | Speech-Lang. Therapy | 16 | 10 | 16 | 1177 | 73.6 |
| Heidlage[178] | 2020 | Speech-Lang. Therapy | 14 | 9 | 21 | 997 | 64.6 |
| Hislop[187] | 2020 | Physiotherapy | 5 | 2 | 8 | 269 | 49.3 |
| Howlett[184] | 2019 | Phys Act/Exercise Int. | 15 | 9 | 15 | 2747 | 195.7 |
| Kuntsler[188] | 2018 | Physiotherapy | 5 | 1 | 5 | 1081 | 216.2 |
| Lim[183] | 2019 | Phys Act/Exercise Int. | 20 | 10 | 31 | 2395 | 107.5 |
| McMichan[468] | 2018 | Phys Act/Exercise Int. | 5 | 2 | 8 | 3569 | 569.8 |
| Nye[179] | 2013 | Speech-Lang. Therapy | 8 | 4 | 8 | 295 | 36.9 |
| Salamh[190] | 2019 | Physiotherapy | 6 | 2 | 15 | 285 | 45.6 |
| Salamh[191] | 2017 | Physiotherapy | 5 | 4 | 6 | 305 | 66.2 |
| Scott[185] | 2018 | Phys Act/Exercise Int. | 20 | 11 | 21 | 2620 | 124.8 |
| Ward [186] | 2020 | Physiotherapy | 12 | 5 | 12 | 806 | 64.7 |
| Young[469] | 2018 | Physiotherapy | 5 | 2 | 6 | 133 | 22.2 |
| 19 | | | 204 | 98 | 249 | 25056 | 118.03 |

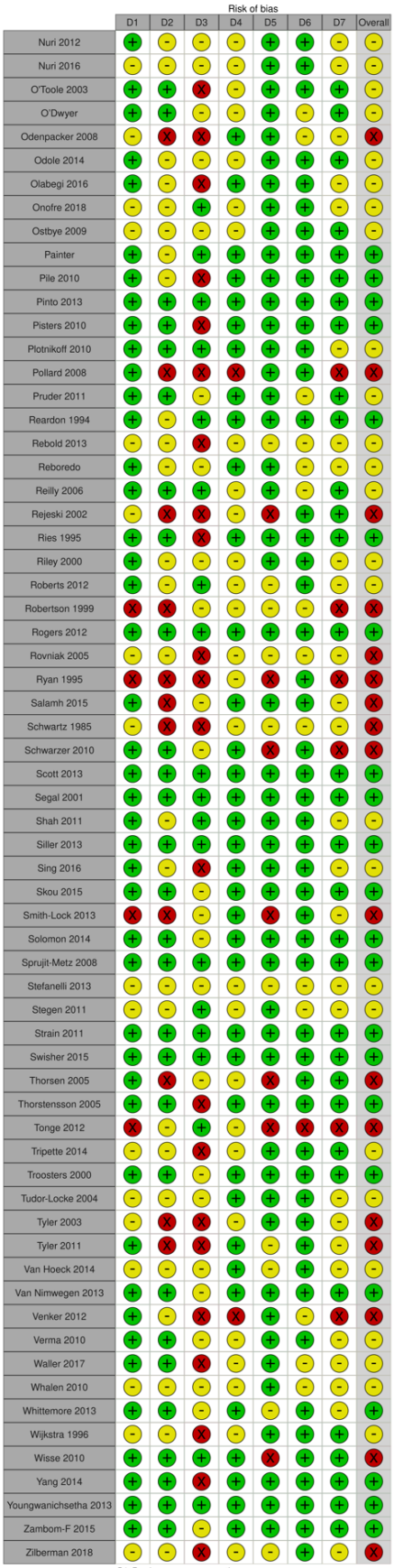
**Table 4.1:** Meta-epidemiological sample- Included meta-analyses and their characteristics.

## Risk of bias of RCTs in the sample of meta-analyses

Two-hundred four studies were assessed with the Cochrane ROB tool. Of these, 73 were scored as having a low risk of bias, 84 had some concerns, and 47 were at high risk (figure 4.3).


Risk of bias
D1 D2 D3 D4 D5 D6 D7 Overall
Drew 2002
Dritsa 2009
Dronkers 2010
Duruturk 2016
Dwyer 2015
Ebenezar 2011
Edvardsen 2015
Eliakim 2007
Emery 1998
Fey 1993
Finch 2014
Fitzgibbon 2010
Fjeldsoe 2010
Franken 2005
Fung 2005
Gallagher 2009
Ghaffari 2012
Giallauria 2016
Giralometto 1996
Goods 2013
Gram 2010
Green 2010
Groussard
Guttentag 2014
Hamar 1990
Hardan 2014
Hardan 2015
Harris 2002
Harrison 2004
Hassanejad 2017
Herring 2017
Hertogh 2010
Holmes 2018
Hornsby 2014
Huang 2011
Huebner 2000
Hunt 2013
Hvid 2016
Jassil 2015
Jones 2005
Jones 2011
Jones 2014
Kasari 2008
Keller 2014
Kim 2006
Koh
Kolt 2006
Lake 1990
Larson 1999
Lattermann 2008
Laudner 2014
Lee 2009
Leermakers 1998
Lewis 2013
Ligtenberg 1995
Lioret 2012
Lonigan 1998
Lovelady 2000
Lovelady 2009
Marconcin 2018
Mehnert 2011
Menske 2010
Michaud 1995
Mohr 2014
Moore 2011
Moreira 2013
Napolitano 2006
Nicklas 2014
Norton 2011


Risk of bias
D1 D2 D3 D4 D5 D6 D7 Overall
Adams 2017
Aglamis 2008
Aittasalo 2012
Al-Majid 2015
Aldred 2004
Alhassan 2007
Alhassan 2012
Alhassan 2013
Anessi 2013
Bailey 2017
Balducci 2010a
Balducci 2010b
Basler 2007
Bauman 2012
Belanger 2013
BelLows 2013
Bennell 2014
Bento 2015
Bertz 2015
Bickmore 2013
Bock 2001
Bonvin 2013
Borghi-Silva 2009
Borghi-Silva 2015
Boyce 2010
Broderick 2014 n
Broman 2006
Brosseau 2012
Buman 2011
Buschman 2009
Buszewicz 2006 n
Cadmus 2010
Campanha 2017
Cardon 2009
Carmack 1995
Carter 2011
Casenhiser 2013
Castelo 2011
Chaipinyo 2009
Chen 1998
Cheung 2009
Coen 2015
Coleman 2012 n
Colleran 2012
Courneya 2008
Courneya 2009
Craig 1996
Craigie 2011
Crane-Thoreson 1999
Crossley 2016
Cui 2018
Dailey 2015
DalLow 2003
Daniels 2017
Davenport 2011
Dawson 2010
De Bock 2013
De Craemer
De Greef 2010
De Greef 2011
De Luca 2016
De Vries 2015
deRosset 2013
Deyle 2000
Deyle 2005
Do 2015
Doi 2008


Risk of bias
D1 D2 D3 D4 D5 D6 D7 Overall
Nuri 2012
Nuri 2016
O'Toole 2003
O'Dwyer
Odenpacker 2008
Odole 2014
Olabegi 2016
Onofre 2018
Ostbye 2009
Painter
Pile 2010
Pinto 2013
Pisters 2010
Plotnikoff 2010
Pollard 2008
Pruder 2011
Reardon 1994
Rebold 2013
Reboredo
Reilly 2006
Rejeski 2002
Ries 1995
Riley 2000
Roberts 2012
Robertson 1999
Rogers 2012
Rovniak 2005
Ryan 1995
Salamh 2015
Schwartz 1985
Schwarzer 2010
Scott 2013
Segal 2001
Shah 2011
Siller 2013
Sing 2016
Skou 2015
Smith-Lock 2013
Solomon 2014
Sprujit-Metz 2008
Stefanelli 2013
Stegen 2011
Strain 2011
Swisher 2015
Thorsen 2005
Thorstensson 2005
Tonge 2012
Tripette 2014
Troosters 2000
Tudor-Locke 2004
Tyler 2003
Tyler 2011
Van Hoeck 2014
Van Nimwegen 2013
Venker 2012
Verma 2010
Waller 2017
Whalen 2010
Whittemore 2013
Wijkstra 1996
Wisse 2010
Yang 2014
Youngwanichsetha 2013
Zambom-F 2015
Zilberman 2018

Risk of bias judgments key:

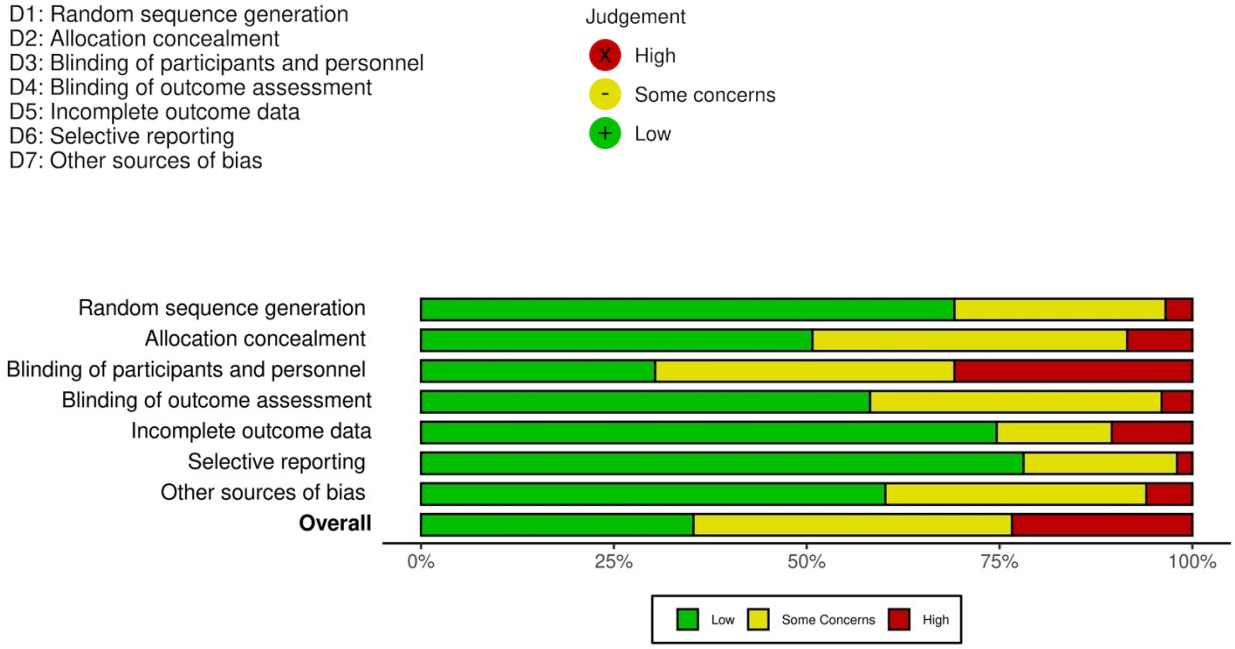



**Figure 4.3:** Risk of bias assessment results, Cochrane Risk of Bias (ROB) tool.

The highest percentages of risk of bias were found in "blinding of participants and personnel" and incomplete outcome data (figure 4.3).

**Objective II:** To quantify and characterize intervention fidelity monitoring and reporting in RCTs in eligible meta-analyses. (AIM I)

## Intervention reporting in the sample of meta-analyses

Of 204 included RCTS, 95, or 46.6 % did not report at least one key aspects of the intervention materials, dosage, intensity, procedures, or the provider qualification or training when assessed with the CONSIDER checklist items 1-5 (figure 4.1). Neary a third (31.9%) of studies failed to fully report key aspects on intervention dosage (frequency, duration, intensity) sufficiently for

reproducibility (did not report or only partially reported). Nearly one quarter (24%) of RCTs publications only partially described the intervention with enough detail to allow for replication. Only 39.7% reported planned or allowed tailoring or modification of the intervention during the trial (table 4.2).

| CONSIDER Checklist item | n (%) Score 0: no report | n (%) Score 1: partial report | n (%) Score 2: mostly reported |
|---|---|---|---|
| 1: Training or intervention materials used in trial | 17 (8.3) | 39 (19) | 148 (72.6) |
| 2: Description of the experimental intervention | 3 (0.02) | 49 (24) | 152 (74.5) |
| 3: Tailoring of intervention (planned or allowed) | 46 (22.5) | 77 (37.8) | 81 (39.7) |
| 4: Intervention providers: qualifications, expertise, or training in intervention or trial protocol received | 71 (34.8) | 61 (29.9) | 72 (35.3) |
| 5: Intervention dose: frequency, duration, intensity | 2 (0.98%) | 63 (31.02) | 139 (68%) |

**Table 4.2:** Reporting of CONSIDER items 1-5

## Fidelity assessment with the CONSIDER checklist:

Among the 204 included RCTs, 137 monitored any degree of participant adherence (frequency of participants' interaction with the intervention), or intervention fidelity for a fidelity monitoring prevalence of 67.2%. Eighty, or 39.2% of trials reported achieving participant adherence of ≥75% during the trial, while 82, or 40.2% of trials achieved greater than partial intervention fidelity among eligible RCTs. Only 35, or 17.2% of RCTs achieved or reported both adherence and fidelity (figure 4.4).

Only 57 (27%) RCT publications described the modifications or tailoring of interventions that took place during the trial. These modifications were reported to be unplanned in nearly a third (18) of those.

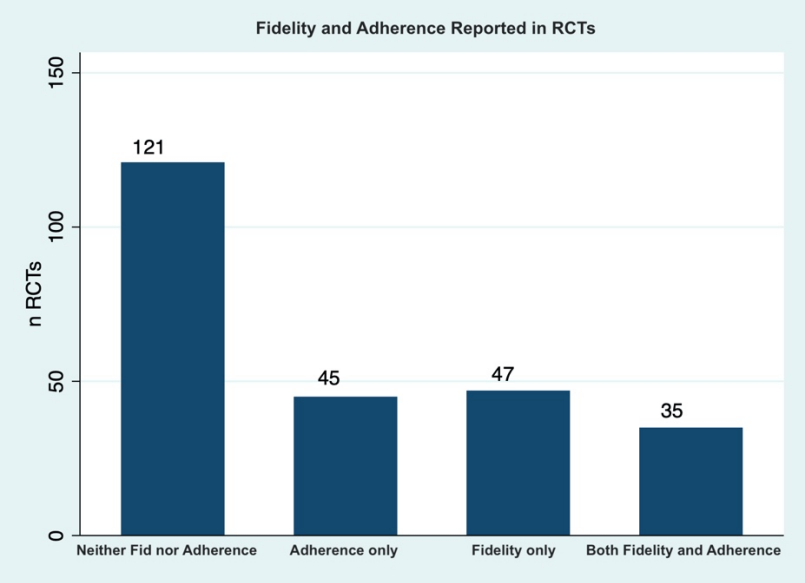



**Figure 4.4:** Fidelity and adherence monitoring and achievement in eligible RCTs

## Fidelity reporting: Was fidelity reported more often in RCTs with smaller treatment effects?

Randomised controlled trials with moderate to large effect sizes reported monitoring fidelity less often than trials with small effect sizes. Thirteen percent more RCTs with small or no treatment effect size reported monitoring fidelity than those with moderate to large effects (table 4.3). Similarly, 14% more RCTs with small treatment effects reported achieving any degree of fidelity RCTs with large treatment effects. A smaller difference was found between RCTs favouring either the control or the experimental intervention and reporting fidelity of monitoring or maintaining any degree of intervention fidelity. Both were reported in nearly equivalent proportions among RCTS favouring either the control or the experimental interventions.

| RCT characteristic | % reporting fidelity only or fidelity plus adherence monitoring or assessment n (%) | % reporting achieving or maintaining fidelity (any degree) n (%) |
|---|---|---|
| RCTs with SMD < 0.50 | 81/128 (63%) | 72/128 (56%) |
| RCTs with SMD ≥ 0.50 | 38/76 (50%) | 32/76 (42%) |
| RCTs with outcome favouring control, or no statistically significant difference experimental/control interventions | 31/53 (59%) | 27/53 (51%) |
| RCTs with outcome favouring experimental intervention | 85/151 (56%) | 75/151(49%) |

**Table 4.3:** Fidelity reporting by trial outcome

**Objective III.** To evaluate the influence of intervention fidelity monitoring on estimates of treatment effects in meta-analysed RCTs in a paired-design, meta-epidemiological (meta-meta-analytic) study. (AIM II)

## Analysing meta-analyses: Differences between RCTs with, and without, intervention fidelity

Within meta-analyses, pooled treatment effect estimates calculated for studies meeting fidelity criteria ranged from SMD (-0.02)[464] to 0.87[179], and 0.13[468] to 0.99[180] for studies without intervention fidelity (table 4.4).

| Meta-Analysis | RCTs: Fidelity | Pooled SMD (95% CI:) | Heterogeneity $I^2$ / $T^2$ | RCTs: No Fidelity | Pooled SMD | Heterogeneity $I^2$ / $T^2$ |
|---|---|---|---|---|---|---|
| Avery 2012[464] | 4 | 0.02 (-0.26,0.30) | 0% / 0.00 | 5 | 0.35 (0.05, 0.66) | 34% / 0.06 |
| Bellicha 2018[465] | 2 | 0.43 (-0.04, 0.91) | 67% / 0.14 | 8 | 0.30 (-0.10, 0.69) | 55% / 0.22 |
| Briani 2018 [466] | 7 | 0.37 (0.20, 0.54) | 38% / 0.02 | 8 | 0.38 (0.08, 0.67) | 80% / 0.12 |
| Cleave 2015[180] | 4 | 0.83 (0.42, 1.23) | 0% / 0.00 | 3 | 0.99 (0.13, 1.85) | 0 % / 0.00 |
| Depiazzi 2019[189] | 5 | 0.43 (0.18, 0.68) | 22% / 0.02 | 2 | 0.86 (0.34, 1.38) | 0% / 0.00 |
| Finch 2016[467] | 6 | 0.42 (-0.29, 1.14) | 96% / 0.73 | 15 | 0.41 (0.16, 0.66) | 87% / 0.19 |
| Hampton 2016[398] | 11 | 0.21 (0.06, 0.36) | 15% / 0.01 | 5 | 0.43 (0.13, 0.73) | 0% / 0.00 |
| Heidlage 2020[178] | 9 | 0.20 (0.06, 0.33) | 10% / * | 5 | 0.56 (0.28, 0.83) | 28% / * |
| Hislop 2020[187] | 2 | -0.02 (-0.35, 0.30) | 36% / 0.04 | 3 | 0.67 (-0.57, 1.90) | 91% / 1.44 |
| Howlett 2019[184] | 9 | 0.19 (0.07, 0.31) | 0% / 0.00 | 6 | 0.46 (0.17, 0.75) | 65% / 0.08 |
| Kuntsler 2018[188] | 2 | 0.19 (-0.04, 0.43) | 0 %/ 0.00 | 3 | 0.18 (-0.21, 0.57) | 81% / 0.09 |
| Lim 2019[183] | 9 | 0.56 (0.26, 0.87) | 80% /0.21 | 11 | 0.61 (0.29, 0.93) | 82% / 0.28 |
| McMichan 2018[468] | 2 | 0.06 (-0.02, 0.14) | 0% / 0.00 | 3 | 0.13 (0.02, 0.23) | 77% |
| Nye 2013[179] | 4 | 0.87 (0.20, 1.55) | 67% / 0.31 | 4 | 0.81 (0.21, 1.40) | 34% / 0.12 |
| Salamh 2019[190] | 3 | 0.38 (0.02, 0.74) | 67% / 0.18 | 5 | 0.70 (0.45, 0.96) | 56% / 0.11 |
| Salamh 2017[191] | 4 | 0.60 (0.40, 0.80) | 58% | 1 | 0.73 (0.10, 1.36) | n/a |
| Scott 2018[185] | 11 | 0.31 (0.12, 0.50) | 21% / 0.02 | 10 | 0.47 (0.30, 0.65) | 0% / 0.00 |
| Ward 2020[186] | 5 | 0.43 (0.04, 0.81) | 0% / 0.00 | 8 | 0.49 (0.29, 0.68) | 0% / 0.00 |
| Young 2018[469] | 2 | 0.51 (-0.13, 1.14) | 6% / 0.01 | 3 | 0.41 (-0.24, 1.06) | 55% / 0.24 |

**Table 4.4:** Pooled SMD and heterogeneity by fidelity

Between meta-analyses, pooled treatment effect sizes were smaller and more precise when fidelity was achieved than when they were absent (figure 4.4). Pooled treatment effect sizes for studies not achieving fidelity or adherence were 0.46 (0.37, 0.55), compared to 0.34 (0.27, 0.41) for studies achieving any combination of adherence or fidelity. As degrees of fidelity increased, pooled treatment effect sizes became smaller and more precise: pooled SMD of 0.53 (0.36, 0.70)

for studies in which only participant adherence was achieved, 0.29 (0.18, 0.40) when intervention fidelity was achieved (but not adherence), and 0.14 (0.18, 0.20) when both fidelity and adherence were achieved together (figure 4.5).

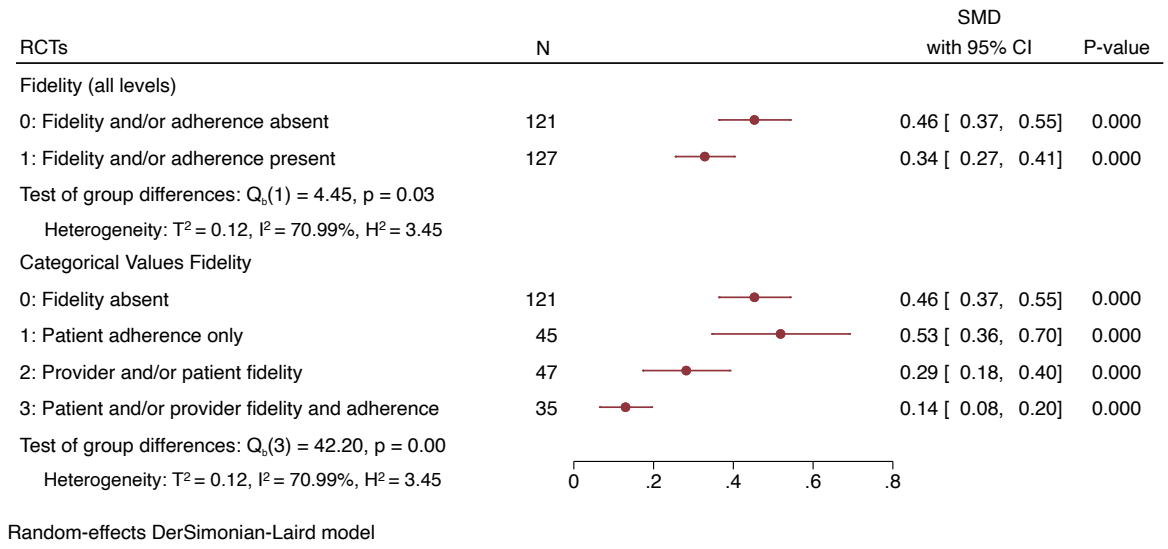



**Figure 4.5:** Pooled treatment effects and degree of fidelity

## CONSIDER Items 1-6 (processes supporting fidelity) and pooled SMDs

The CONSIDER checklist was designed to assess fidelity multidimensionally, given the wide range of terms and constructs for fidelity and its components found, and often used interchangeably, in complex intervention literature. The checklist's items 1-6 assess processes supporting intervention fidelity, while items 7-8 assess fidelity directly. Analysis of the results of this assessment found that smaller and more precise pooled effect estimates were also found for trials scoring 9 or greater, or supporting fidelity, on the CONSIDER checklist than those scoring less than 9, or not supporting intervention fidelity (table 4.5).

| Fidelity process score category | N RCTs | Pooled SMD (95% CI:) | Heterogeneity $I^2$ / $T^2$ |
|---|---|---|---|
| 1. Consider checklist ≤ 6 | 60 | 0.40 (0.28, 0.52) | 70.4%/0.12 |
| 2. Consider checklist ≥ 9 | 101 | 0.27 (0.21, 0.33) | 67.0%/ 0.10 |
| Overall p <0.00001 | | | 67.5%/0.08 |

**Table 4.5:** Pooled SMDs: CONSIDER scores supporting fidelity

## Meta-epidemiological analyses: Does poor intervention fidelity bias treatment effect estimates?

To investigate the impact of monitoring or maintaining fidelity on RCTs' treatment effect estimates, differences in SMDs (dSMD) between trials with and without fidelity were derived for each meta-analysis and pooled in a meta-meta-analysis (Figure 4.6). The resulting median difference was 0.13 (95% CI: 0.07,0.20) in favour of RCTs not maintaining fidelity, with low evidence for variability in dSMDs across meta-analyses ($t^2$ 0.01) above of what would be expected by chance. This means that RCTs not maintaining intervention fidelity had larger and less precise (0.13 95%CI:-0.30, 0.56] SMDs between the experimental and control interventions than RCTs in which intervention fidelity was maintained.

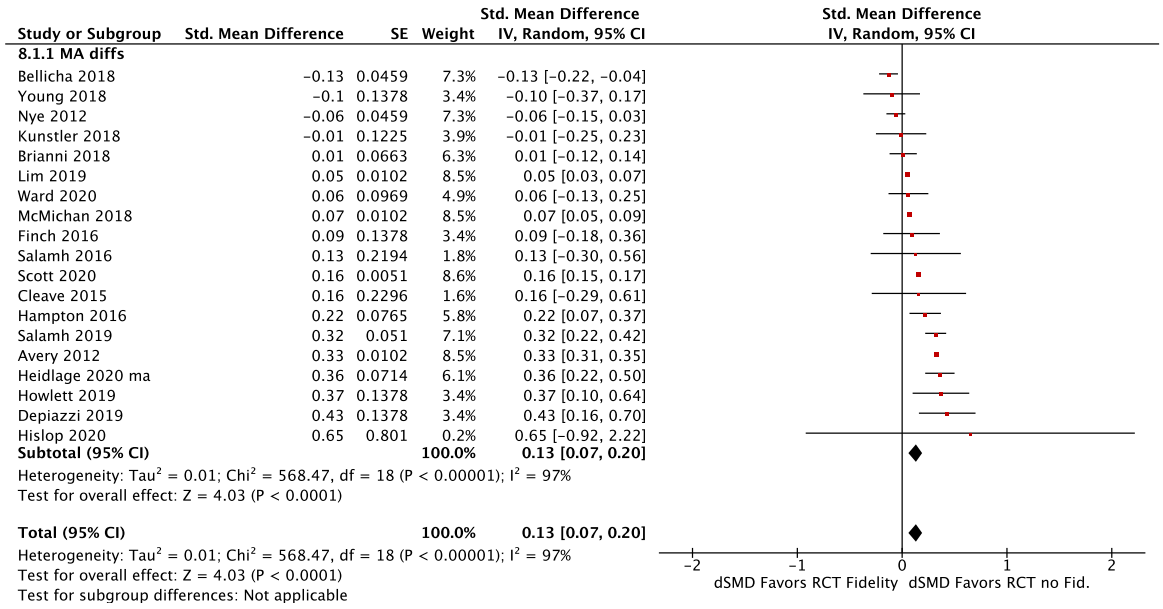



**Figure 4.6:** dSMD in meta-analyses with fidelity and adherence and RCTS without fidelity or adherence.

**Objective IV:** To determine the direction of the effect of poor intervention fidelity on treatment effect estimates in meta-analysed RCTs in AIM II.

## Meta-regression:

In this section, we investigated how much treatment effect estimates may change as intervention fidelity increases or decreases. Random effects meta-regression (DerSimonian-Laird) identified a linear relationship between fidelity and treatment effect sizes across RCTs (figure 4.7).

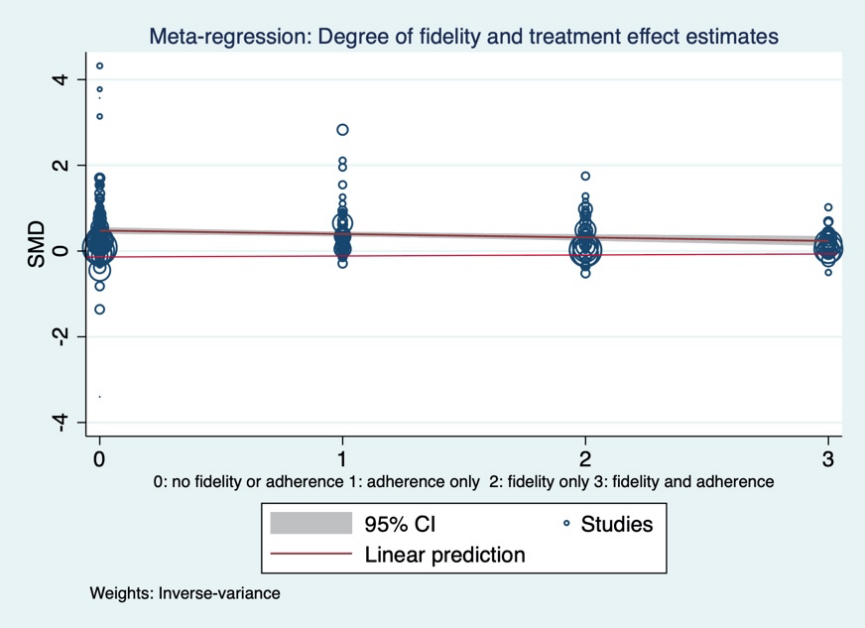



**Figure 4.7:** Bubble plot, Fidelity (categorical) and treatment effect size

Treatment effects sizes changed by (-0.23) (95% CI: -0.38,-0.74) when both intervention fidelity and participant adherence, measured as a binary outcome, were maintained in RCTs (table 4.6). Adjusting for both risk of bias (low or unclear and high combined) and sample size did not significantly alter the relationship between intervention fidelity and treatment effect estimates, -0.22 (95% CI: -0.07,-0.36).

| | **Meta-regression: All RCTs** | **Coef.** | **Std. Error** | **Z** | **p** | **95% CI:** | **$I^2$ %** | **$T^2$** | **$R^2$ %** |
|---|---|---|---|---|---|---|---|---|---|
| 1. | Fidelity and adherence (binary variable) | -0.231 | 0.080 | -2.88 | 0.004 | -0.39, -0.736 | 70.64 | 0.12 | 0.84 |
| a | Fidelity and adherence, sample size (100 person units), risk of bias (low, unclear+high) | -0.215 | 0.075 | -2.86 | 0.004 | -0.363,-0.068 | 65.55 | 0.95 | 19.67 |
| | Sample size | -0.088 | 0.016 | -5.51 | 0.000 | -0.119, -0.057 | 65.55 | 0.95 | 19.67 |
| | Risk of bias | -0.055 | 0.565 | -0.98 | 0.327 | -0.166, | 0.055 | 0.95 | 19.74 |
| 2. | Fidelity categorical: 0 none, 1 adherence, 2 fidelity only ,3 fidelity + adherence | -0.081 | 0.026 | -3.15 | 0.002 | -0.131, -0.031 | 70.53 | 0.12 | 1.30 |
| a | Fidelity categorical, sample size (100 person units), risk of bias (low, unclear + high) | -0.073 | 0.024 | -2.99 | 0.003 | -0.123, -0.025 | 65.73 | 0.95 | 19.01 |
| | Sample size | -0.088 | 0.016 | -5.46 | 0.000 | -0.119, -0.056 | 65.73 | 0.95 | 19.01 |
| | Risk of bias | -0.043 | 0.064 | -0.64 | 0.502 | -0.167, 0.082 | 65.73 | 0.95 | 19.01 |

**Table 4.6:** Meta-regression coefficients

The impact of fidelity as a categorical value was also assessed by meta-regression as:

0: fidelity absent,

1: only participant adherence present,

2: fidelity present

3: fidelity and participant adherence present.

Treatment effect sizes (SMD) changed by (-0.08) (95% CI: -0.13,-0.31) for every increase in fidelity category. Including both risk of bias as a categorical variable (low, high, unclear) and study sample size (categorical variable in 100 participant units) in the meta-regression found that SMDs changed by (-0.07) (95% CI:-0.03, -0.12) per unit increase in fidelity when adjusting for sample size and risk of bias (table 4.6).

### Subgroup analyses:

This section explores how various factors influence the relationship between intervention fidelity and treatment effect estimates.

### Heterogeneity

Between-trial heterogeneity was lowest for pooled treatment effects among RCTs meeting fidelity and participant adherence criteria, with $I^2$ of 0%, rising as levels of fidelity decreased to $I^2$ of 60%, when fidelity without participant adherence was achieved, $I^2$ of 77% when only adherence was achieved, and 73% when neither adherence nor fidelity were achieved (table 4.7). See appendix IV figure 4.2 for forest plots of pooled meta-analyses.

| Level of Fidelity | N comparisons | Pooled SMD (95% CI:) | Heterogeneity $I^2$ / $T^2$ |
|---|---|---|---|
| 0: No Fidelity or Adherence (Consider checklist) | 121 | 0.47 (0.38, 0.56) | 74% / 0.15 |
| 1: Adherence only (participant and/or provider) | 45 | 0.53 (0.36, 0.70) | 76.51 % / 0.41 |
| 2: Intervention fidelity only (participant and/or provider) | 47 | 0.29 (0.18, 0.40) | 61.54% / 0.06 |
| 3: Fidelity and adherence (participant and/or provider) | 35 | 0.14 (0.08, 0.20) | 0 % / 0.00 |

**Table 4.7:** Heterogeneity

## Risk of bias

Subgroup analysis using only RCTs judged as low risk of bias on the ROB-2 found that pooled treatment effect estimates decreased in size and increased in precision with increasing levels of fidelity (figure 4.8), following the same pattern that was observed in the main analyses. Meta-regression identified the same linear relationship between fidelity and treatment effect sizes across RCTs (figure 4.9) and a nearly identical regression coefficients for combined fidelity and adherence, including when adjusting for sample sizes (table 4.8).

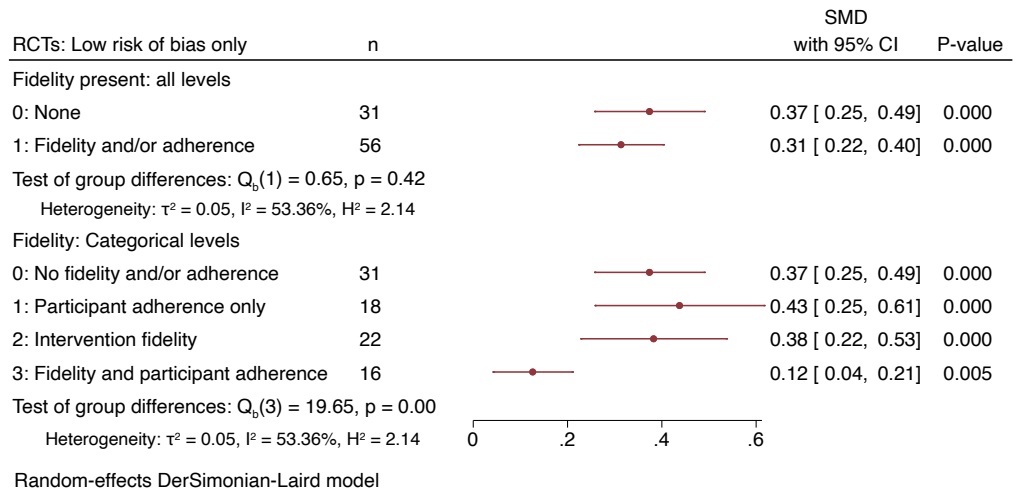



**Figure 4.8:** Subgroup analyses, risk of bias (Cochrane ROB-2).

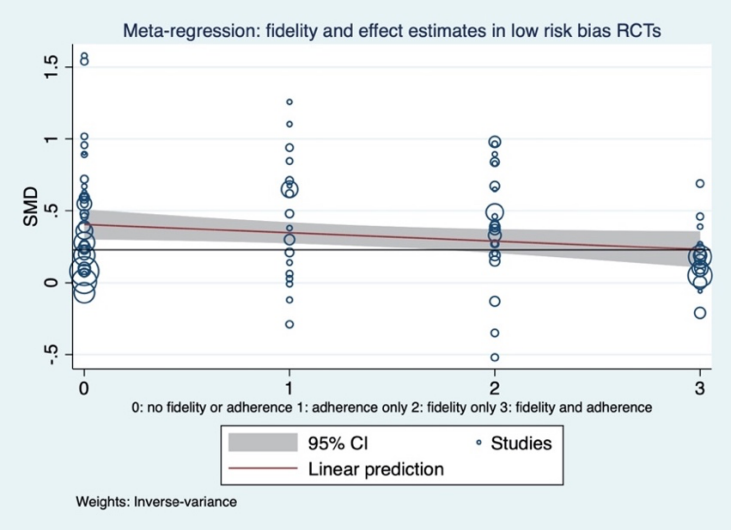



**Figure 4.9:** Bubble plot, fidelity (categorical) and treatment effect size among RCTs with low risk of bias

| | Meta-regression: RCTs with Low risk of bias | Coef. | Std. Error | Z | p | 95% CI: | $I^2$ % | $T^2$ | $R^2$ % |
|---|---|---|---|---|---|---|---|---|---|
| 1. | Fidelity and adherence (binary variable) | -0.235 | 0.085 | -2.75 | 0.006 | -0.675,- 0.402 | 50.56 | 0.42 | 8.48 |
| a | Fidelity and adherence, sample size (categorical, 100 person units) | -0.201 | 0.078 | -2.56 | 0.01 | -0.355, -0.047 | 40.7 | 0.29 | 35.79 |
| | Sample size | -0.053 | 0.017 | -3.01 | 0.003 | -0.185, -0.088 | 40.7 | 0.29 | 35.79 |
| 2. | Fidelity categorical: 0 none, 1 adherence, 2 fidelity only ,3 fidelity + adherence | -0.058 | 0.031 | -1.90 | 0.05 | -0.118, -0.002 | 53.59 | 0.05 | 0 |
| a | Fidelity categorical, sample size (100 person units | -0.061 | 0.027 | -2.32 | 0.020 | -0.119, -0.01 | 42.55 | 0.03 | 30.69 |
| | Sample size | -0.064 | 0.018 | -3.53 | 0.000 | -0.1, -0.029 | 42.55 | 0.03 | 30.69 |

**Table 4.8:** Fidelity meta-regression coefficients

## Sample size

Treatment effect sizes were larger in trials with small sample sizes than trials with more than 100 participants (figure 4.10). In both trials with less than or more than 100 participants, pooled treatment effect sizes decreased, and precision increased as levels of fidelity increased (figures 4.10a and 4.10b).

The greatest difference was found between trials with absent fidelity or adherence in both trials with large sample sizes (SMD 0.24, 95% CI:0.13,0.34) and small sample sizes (SMD 0.63, 95% CI:0.49,0.76) and trials with patient and provider fidelity and adherence (SMD 0.10, 95% CI:0.03,0.18 large trials, SMD 0.22, 95% CI: 0.11,0.34 small trials). Measures of heterogeneity ($I^2$) and between trial variance ($T^2$) were 0.00 for trials with the highest level of fidelity, category 3, in large and small trials, and considerably higher when fidelity was absent in both large ($I^2$ 76.73, $T^2$ 0.076) and small sample sized trials ($I^2$ 63.84, $T^2$ 0.218).

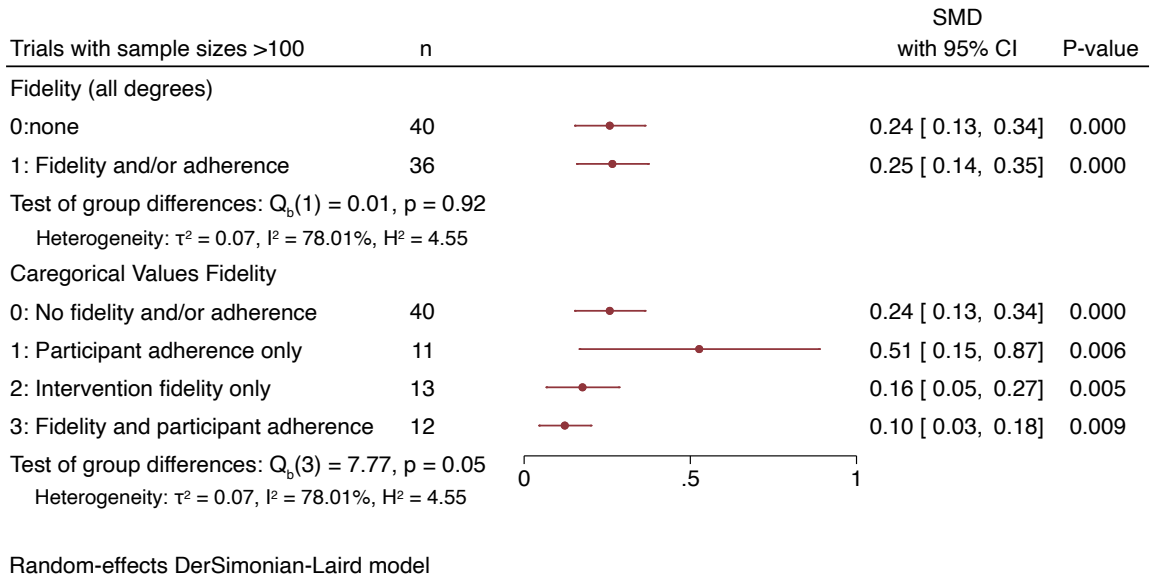



**Figure 4.10a:** Subgroup analysis, RCT sample size >100

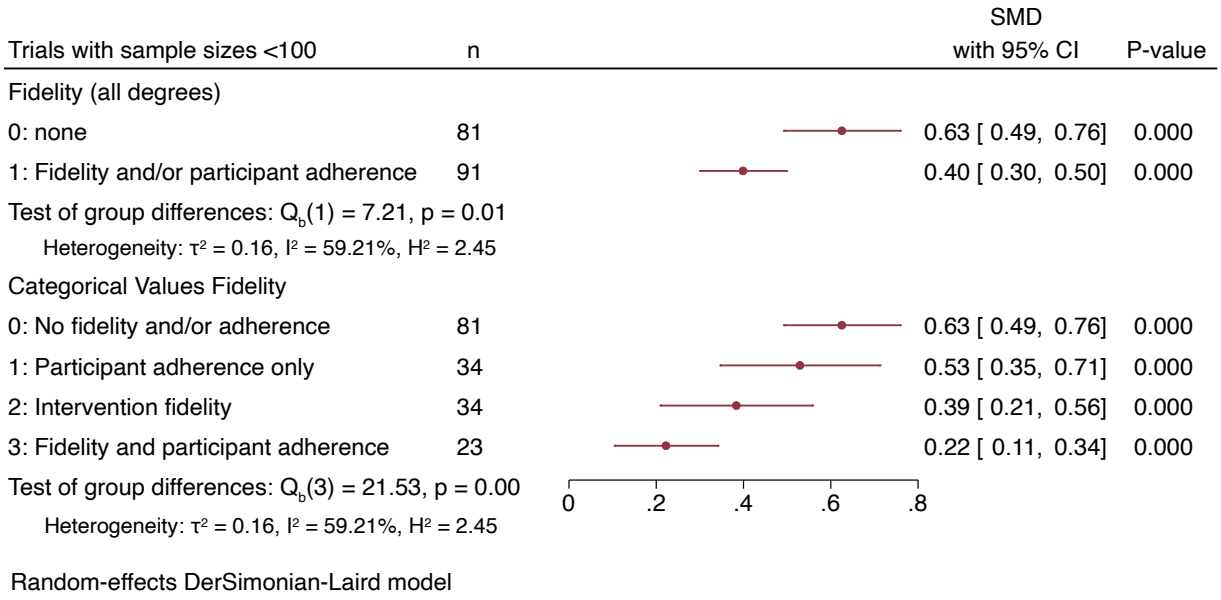



**Figure 4.10b:** Subgroup analysis, RCT sample size <100

## Intervention fidelity and meta-analyses: How would poor intervention fidelity impact the results of meta-analyses?

In this section, we estimate how the results of the sample of meta-analyses used in the meta-epidemiological study would change if the fidelity meta-regression coefficient (0.23) was applied to RCTs not reporting or maintaining intervention fidelity. This allowed us to simulate what impact poor or absent intervention fidelity could have on the results of meta-analysis of rehabilitation RCTs. Each meta-analysis was then recalculated (generic inverse variance, random effects meta-analysis) with the adjusted SMDs (labelled as adjusted pooled effect in table 4.9) and compared to the pooled results reported in the published meta-analyses (labelled "unadjusted pooled effect" in table 4.9).

Comparison of pooled treatment effect estimates from meta-analyses before and after adjustment with the corresponding fidelity meta-regression coefficients from table 4.9 resulted in decreased treatment effect sizes (SMD) in 19 out of included meta-analyses (table 4.9). The statistical significance of pooled treatment effects was lost in 4 of 19 meta-analyses.[464,465,467,469] Adjusted confidence intervals were more narrow (increased precision) in 14 of 19 meta-analyses after adjusting for fidelity.

Additionally, the magnitude of the treatment effect, as categorized by Cohen[457], decreased in 4 of 10 meta-analyses.[183,184,464,465] In the meta-analyses by Bellicha[465] and Lim[183], the treatment effect size decreased from moderate to small, while in treatment effect sizes for Avery[464] and Howlett[184] decreased from small to nearly no effect, or 0.06 (95% CI:: -0.12,0.25) and 0.03 (95% CI:: 0.01, 0.26), respectively.

| Meta-analysis | Pooled effect | $I^2$% / $t^2$ | Pooled SMD ( 95% CI:) | Direction of change: SMD | Change stat. significance | Change in Cohen's effect category |
|---|---|---|---|---|---|---|
| **Avery 2012** [464] | Unadjusted<br>Adjusted | 0% 0.00<br>19% 0.03 | 0.22 (0.02, 0.41)<br>0.06 (-0.12, 0.25) | ↓ | ✓ | ✓ small to none |
| **Bellicha 2018** [465] | Unadjusted<br>Adjusted | 58% 0.15<br>58% 0.15 | 0.39 (0.03. 0.74)<br>0.17 (-0.18, 0.52) | ↓ | ✓ | ✓ mod to small |
| **Briani 2018** [466] | Unadjusted<br>Adjusted | 72% 0.07<br>66% 0.05 | 0.35 (0.18, 0.53)<br>0.21 (0.05, 0.37) | ↓ | x | x |
| **Cleave 2015** [180] | Unadjusted<br>Adjusted | 0% 0%<br>0% 0% | 0.86 (0.49, 1.22)<br>0.78 (0.41, 1.14) | ↓ | x | x |
| **Depiazzi 2019** [189] | Unadjusted<br>Adjusted | 22% 0.03<br>0% 0% | 0.50 (0.27, 0.74)<br>0.40 (0.20, 0.59) | ↓ | x | x |
| **Finch 2016** [467] | Unadjusted<br>Adjusted | 93% 0.34<br>93% 0.34 | 0.44 (0.11, 0.75)<br>0.29 (-0.01, 0.60) | ↓ | ✓ | x |
| **Hampton 2016** [398] | Unadjusted<br>Adjusted | 0% 0<br>0% 0 | 0.26 (0.14, 0.38)<br>0.19 (0.07, 0.31) | ↓ | x | x |
| **Heidlage 2020** [178] | Unadjusted<br>Adjusted | 30% 0.04<br>13% 0.01 | 0.33 (0.17, 0.49)<br>0.23 (0.10, 0.37) | ↓ | x | x |
| **Hislop 2020** [187] | Unadjusted<br>Adjusted | 87% 0.63<br>87% 0.61 | 0.25 (-0.35, 0.84)<br>0.17 (-0.24, 0.57) | ↓ | x | x |
| **Howlett 2019** [184] | Unadjusted<br>Adjusted | 35% 0.02<br>34% 0.02 | 0.27 (0.15, 0.39)<br>0.03 (0.01, 0.26) | ↓ | x | ✓ small to none |
| **Kuntsler 2018** [188] | Unadjusted<br>Adjusted | 63% 0.04<br>74% 0.06 | 0.19 (-0.04, 0.41)<br>0.12 (-0.05, 0.28) | ↓ | x | x |
| **Lim 2019** [183] | Unadjusted<br>Adjusted | 81% 0.23<br>80% 0.22 | 0.62 (0.35, 0.91)<br>0.46 (0.18. 0.75) | ↓ | x | ✓ mod to small |
| **McMichan 2018**[468] | Unadjusted<br>Adjusted | 57% 0.01<br>63% 0.02 | 0.08 (-0.2, 0.18)<br>0.02 (-0.06, 0.10) | ↓ | x | x |
| **Nye 2013** [179] | Unadjusted<br>Adjusted | 50% 0.18<br>45% 0.22 | 0.86 (0.43, 1.28)<br>0.72 (0.28, 1.16) | ↓ | x | x |
| **Salamh 2019** [190] | Unadjusted<br>Adjusted | 48% 0.04<br>49% 0.05 | 0.62 (0.35, 0.88)<br>0.53 (0.26, 0.79) | ↓ | x | x |
| **Salamh 2017** [191] | Unadjusted<br>Adjusted | 69% 0.20<br>67% 0.19 | 0.74 (0.41, 1.06)<br>0.58 (0.28, 0.88) | ↓ | x | x |
| **Scott 2018** [185] | Unadjusted<br>Adjusted | 4% 0<br>26% 0.03 | 0.38 (0.28. 0.61)<br>0.19 (0.05, 0.33) | ↓ | x | x |
| **Ward 2020** [186] | Unadjusted<br>Adjusted | 0% 0<br>0% 0 | 0.47 (0.29, 0.65)<br>0.25 (0.07, 0.42) | ↓ | x | x |
| **Young 2018** [469] | Unadjusted<br>Adjusted | 36% 0.11<br>36% 0.11 | 0.44 (0.00, 0.88)<br>0.24 (-0.20, 0.69) | ↓ | ✓ | x |

**Table 4.9:** Pooled SMD for meta-analysis, published (unadjusted) and adjusted with fidelity coefficient if fidelity absent.

## DISCUSSION

In a sample of 19 meta-analyses and 204 RCTs representing the experience of 25,056 participants in physiotherapy, occupational therapy, speech-language therapy, exercise, and physical activity intervention clinical trials, this meta-epidemiological study found evidence of systematic differences between effect estimates in complex intervention trials achieving intervention fidelity compared to those without it. Poor intervention fidelity can significantly influence the interpretation and appropriateness to change practice of complex intervention RCTs.

### Fidelity monitoring in rehabilitation

Fidelity monitoring may improve the internal validity, generalizability and reproducibility of clinical trials.[133] In the absence of fidelity monitoring, it may be difficult to determine if RCTs' results are due to intervention effectiveness, variations in implementation, or unknown or external factors.[133] Yet, this study found that fidelity assessment is not systematic in RCTs of complex interventions in healthcare. Fidelity monitoring within included RCTs in this study was generally poor. Less than a third reported monitoring participant adherence and intervention fidelity while less than half monitored either intervention fidelity, participant adherence, or the frequency of participants' interaction (dosage) with the intervention.

These results support previous findings that fidelity monitoring is poor in clinical trials of complex interventions. In a 2021 systematic review of the characteristics of surgical RCTs published between 2008-2020, Robinson et al. found that only 4.4%, or 17 out of 388 RCTs assessed the

quality of the intervention delivery.[90] Toomey et al. systematically reviewed physiotherapy-led self-management trials of exercise and education[172], finding that only 20-30% of studies reported monitoring key aspects of intervention fidelity (intervention delivery and receipt) when assessed with the National Institutes of Health Behaviour Change Consortium (NIH-BCC) fidelity checklist.[133] Analyses of trials in other complex interventions have also found poor fidelity monitoring and reporting.[50] In an analysis of 202 interventional trials of Mindfulness-based interventions (MBIs) with the NIH-BCC checklist[133], Ketchter et al., that only 25 (12%) described study treatment fidelity.[157] Only half of those (6%), reported fidelity monitoring or data across treatment design, delivery, or receipt, the three key components of the CONSIDER framework. In a systematic review of afterschool intervention programs for at-risk youth, Maynard et al. found that only 29% of studies of measured fidelity and only 4% included fidelity measures in their analyses of the intervention,[470] while Naleppa and Cagle's analysis of 63 social work intervention studies found that only 15% monitored or collected fidelity data.[471]

### Overall fidelity and treatment effect estimates

If intervention fidelity is not monitored and maintained, variations in intervention delivery may bias the results of RCTs.[53,63,89,90] Contrary to our original hypothesis, this meta-epidemiological study of complex intervention RCTs found evidence of larger treatment effect estimates (over-estimation) and wider confidence intervals (decreased precision) in RCTs in which intervention fidelity was not achieved, rather than a reduction in treatment effects. Pooled SMDs within and across meta-analyses were larger and less precise in RCTs in which fidelity was not present or reported. Differences in SMDs between meta-analyses also indicate overestimation of treatment

effect sizes in RCTs in which fidelity was absent across the sample of 19 meta-analyses. Statistical heterogeneity was higher in meta-analyses of trials without intervention fidelity or fidelity and participant adherence, which may reflect variations in the quality of intervention delivery between trials influencing treatment effects.[472]

This was an unexpected finding, and contrary to what we had hypothesized. The mechanisms underlying this finding cannot be determined through meta-analysis and meta-epidemiology. However, it is possible that the direction of bias identified in our meta-epidemiological study reflects the cumulative effects of additional, unplanned interventions and active ingredients introduced by poor intervention fidelity or protocol departures.[57,75,472] When fidelity is not monitored or maintained, there may be greater scope for participant exposure to additional effect-mediating cointerventions or increases in intervention intensity, which may lead to a larger treatment effect, as proposed by Hettema.[473,474]

The links between treatment fidelity and intervention outcomes are complex, however. An intervention could also be delivered with a high degree of fidelity and yet be a poorly conceptualized intervention, or a well conceptualized intervention that is just ineffective. In those scenarios, the association between the degree of intervention fidelity and outcome would be negative, but not because of a high degree of fidelity. Conversely, a robust intervention delivered with poor intervention fidelity may result in large treatment effect estimates, despite, and not because of, poor fidelity. Intervention fidelity is not the only factor moderating treatment outcomes. Fidelity does, however, provide researchers with confidence that an

intervention was tested adequately and that a trial's results are less likely to be due to other potential confounders resulting from poor intervention delivery.[39]

Findings from systematic reviews and meta-analyses of complex interventions with poor fidelity may be unreliable for informing clinical decision making.[436,437] In this study's sample, 128 RCTs had effect sizes (SMDs) for primary outcomes between of less than 0.50. A difference of (0.23) could bias the outcome of a trial from a negligible to a small intervention effect size, a small to moderate effect size, or a moderate to a large effect size. This was also found when comparison pooled SMDs in our sample of meta-analyses before and after adjustment with the fidelity meta-regression coefficient (table 4.9).[475] The pooled treatment effect estimates reduced from moderate to small effect size in two meta-analyses[183,465], and from small to no effect in two others.[184,464] The decrease in precision and loss of statistical significance for treatment effects seen in 4 of the meta-analyses after adjusting for fidelity (table 4.9) also illustrate how poor or absent intervention fidelity can bias the results of systematic reviews and meta-analyses of complex intervention RCTs. This could lead clinicians to implement interventions which may be ineffective or less effective than others or reject potentially effective interventions based on biased data from meta-analyses of RCTs.

Caution is needed when interpreting the results of complex intervention RCTs when fidelity is not monitored or is monitored but not reported.[437] Without monitoring intervention fidelity, the actual intervention dose received by participants cannot be quantified.[476] Poor intervention fidelity may bias the results of RCTs, as indicated by the linear relationship between fidelity and

treatment effect estimates identified with meta-regression in this study. Treatment effect sizes increased by SMD 0.23 (95% CI: 0.04, 0.74) when fidelity was absent, or SMD 0.24 (95% CI: 0.40, 0.68) when adjusted for both RCT sample size and risk of bias. This can have important implications for the interpretation of outcomes from complex intervention RCTs and meta-analyses of RCTs. Similar to our findings, Yoon et al. found that monitoring and maintaining intervention fidelity was associated with smaller, more precise effect sizes in psychological interventions with persons with poor mental health, attributing the finding to intervention fidelity monitoring being associated with more stringent intervention delivery, and less propensity to toward bias during study conduct.[477]

A meta-analysis of Motivational Interviewing (MI), a complex behavioural intervention, by Hettema also found that treatment effect sizes were two times larger in studies without manualized interventions and therapist treatment manuals than those with them.[75] Hettema examined relationships between treatment effect size and a number of study attributes as potential moderators of the outcomes. Of the intervention delivery characteristics examined, only use of a treatment manual during the trial was significantly related to treatment outcomes, predicting 8.5% of the variance treatment effects ($\beta = -0.292$, $p < 0.05$). Studies not reporting use of an in-trial treatment manual had a mean treatment effect size (SMD) of 0.65 (SD = 0.62), whereas those standardizing treatment with a manual reported a SMD of 0.37 (SD = 0.62). The mechanisms underlying this effect could not be investigated in the meta-analysis, however. The findings may reflect an additive effect of additional intervention components introduced by

variations in MI delivery in the non-manualized studies, or differences in tailoring of the intervention to individuals between manualized and non-manualized studies.

Our findings contrast with those of a Ellefson and Oppenheimer's large-scale simulation study of interventions in education.[50] From an initial, computer simulated RCT of an imaginary intervention conducted in 40 classrooms of 25 students each, Ellefson and Oppenheimer generated data for 1000 computer simulated participants across 11,055 trials in 50 computer modelled classrooms and laboratories. Data was analysed across three models including simulated variables for classroom and lab variables such as schools' level of student performance and resource availability. Simulated effect sizes decreased as simulated fidelity decreased, with every 5% fidelity reduction producing a 5% reduction in effect size. Lowering fidelity also increased variance, with greater variance among simulated trials with large effect sizes.

While both our findings and Oppenheimer and Ellison's show that poor fidelity biases treatment effect estimates in complex intervention clinical trials, the direction of the bias is different. The meta-epidemiological analyses in our study found that effect sizes decreased, and their precision increased as the degree of fidelity increased. The difference in the direction of bias may result from the many differences between medical or rehabilitation interventions and interventions in education. The difference in the direction of bias arising from poor fidelity in Ellefson and Oppenheimer's simulations and this meta-epidemiological study may also result from differences in how fidelity was conceived and measured in both studies. Ellefson and Oppenheimer investigated the effects of fidelity on study replicability, not intervention effectiveness, and

modelled fidelity as the percentage of an intervention to which each simulated classroom adhered in the simulation, ranging from (.01) – (0.99), at intervals of .20 for individual classrooms or labs.

However, intervention delivery, or adherence as defined by Ellefson and Oppenheimer, is only one component of fidelity.[40] An intervention can be delivered with the intended frequency or a high percentage of its intended components but be delivered poorly, introducing unplanned, extraneous components and variability.[1,478] Ellefson and Oppenheimer's fidelity construct does not include the quality of intervention delivery (fidelity, as defined in the CONSIDER framework) or the influence of both frequency (adherence) and fidelity combined. Our meta-epidemiological analyses found that the degree of fidelity maintained (adherence and quality of intervention delivery) influenced the magnitude of fidelity's effect on treatment effect estimates in pooled meta-analysis. Compared to RCTs with no fidelity monitoring, or in which fidelity was not maintained, pooled treatment effect estimates were lower when any degree of adherence and fidelity were maintained together. These linear relationships remained when adjusted for study quality and sample size.

The difference in the direction of bias of treatment effect estimates introduced by lower degrees of fidelity in Ellefson and Oppenheimer's study may also be influenced by the difficulty of simulating real-world conditions and treatment effects. Ellefson and Oppenheimer modelled treatment effects, but not treatments. By contrast, our results were derived from meta-epidemiological analyses of the experience of 25,056 participants and their intervention

providers with interventions across 204 physiotherapy, occupational therapy, speech-language therapy, exercise, and physical activity RCTs.

In Ellefson and Oppenheimer's simulation study, no specific intervention or control were simulated. Intervention-specific characteristics such as complexity and context or provider characteristics such as competence and equipoise, can influence both intervention fidelity and treatment effects.[2,479] Other sources of intervention complexity, and potential sources of bias, are difficult to model with computer-simulated data. These include intervention complexity, difficulty maintaining intervention fidelity while tailoring interventions to individual participants, participant responses to interventions or interventionists.

These features would be difficult to model with simulated data but could influence the direction of bias on treatment effect estimates. However, pragmatic RCTs (pRCT), which offer practitioners "considerable leeway in deciding how to formulate and apply"[208] interventions, do mirror real-world conditions and offer an important opportunity to explore the influence of these features and intervention fidelity on intervention effectiveness.[5,7,13,208] Intervention fidelity has been poorly explored in pragmatic RCTs in complex interventions, however. In the following chapter of this thesis (Chapter V), the application of the CONSIDER fidelity framework in an ongoing pragmatic RCT is used to investigate fidelity and adherence and how violations of them affect interpretation and analysis of pRCT results.

## Adherence and dosage

This study also found that pooled treatment effect sizes were often the largest when only participant adherence, or attendance in trial activities, was monitored and achieved, without monitoring or maintenance of intervention fidelity. This was found for SMDs pooled across all meta-analyses, in trials with greater than 100 participants, and among RCTs with unclear or high risk of bias. In RCTs in which only participant adherence was achieved, larger pooled treatment effect estimates may reflect receipt of increased dosage, or increased exposure to, the experimental interventions.

The relationship between adherence and treatment effect estimates in this study supports previous hypotheses that session attendance is an insufficient proxy for fidelity or intervention quality, and may lead to over- or under-estimation of true treatment effects because of unknown exposure to the intervention's active ingredients.[472–474] As with drugs, there is a dose-response effect curve with many complex interventions.[480–482] In the case of physiotherapy or exercise based interventions, for example, participants' physiological responses to interventions are determined by their mode, and dose, or frequency, intensity, and duration, which rely on participant attendance or exposure (adherence) with intervention activities. Without monitoring for both fidelity and adherence, participants may exercise at higher or lower percentages of their maximal heart rate, with less or greater support, or for shorter or longer periods than intended if fidelity is not monitored (and acted upon) in the trial.[472–474]

### What other factors may be at work? sample size, risk of bias, or methodological quality?

Fidelity is complex and may interact with other factors to influence intervention effects in a clinical trial. We explored whether sample size and methodological quality moderate the association between fidelity and treatment effect estimates. The over-estimation of treatment effect estimates we found with poor or absent intervention fidelity also remained in subgroup analysis of RCTs by both sample size and risk of bias. Previous meta-epidemiological studies examining the influence of sample size on the findings of RCTs have found larger treatment effect estimates and decreased precision in RCTs with smaller sample sizes, while larger sample sizes are associated with more accurate mean values, decreased random variability, and greater precision.[79,82,463,483] The treatment effect estimates in our sample were larger (nearly twice or more as large) in trials with small sample sizes (less than 100 participants) than in RCTs with more than 100 participants for all categories of fidelity and adherence (figure 4.6). Our meta-epidemiological analyses also found smaller treatment effect sizes and increased precision in RCTs with large sample sizes (greater than 100 participants), and larger treatment effect estimates and less precision in RCTs with smaller samples (less than 100 participants). This may reflect increased random variability, imprecision, and potential for outliers to skew treatment effects in studies with smaller sample sizes.[463,483] The nearly twice as large treatment effect sizes seen in RCTs with absent fidelity and small sample sizes than in RCTs with absent fidelity and larger samples may reflect a synergistic increase in random variability influenced by small sample size and poor fidelity, rather than small sample size or poor fidelity alone.

The observed association between sample size and treatment effect may also reflect bias from

studies' methodological quality. Previous meta-epidemiological studies have found overestimation of treatment effect estimates in RCTs when risk of bias is unclear or increased as a result of decreased methodological quality.[81,447,462] Sample size may also influence methodical quality, however. Large trials have been associated with higher methodological quality than small trials.[484] Large RCT sample sizes may also reflect the influence of factors including trialists' skill and expertise with recruitment, which may also influence participants' adherence and fidelity to the intervention assigned.[443]

Smaller trials may take place at earlier stages of intervention development, before some intervention parameters are fully defined and tested in larger or multicentre trials.[3,5,7] This may lead to more variable intervention delivery in smaller trials, compared to larger trials conducted at later stages of intervention development. [3,7] Smaller trials are also more likely to be single-centre and have less stringent controls (oversight) than larger trials, potentially contributing to lower methodological quality compared to larger and multicentre studies.[70] Smaller RCTs have also been associated with decreased likelihood of external or public funding, which has been associated with significantly lower methodological quality compared to RCTs without funding.[485]

### Reporting of intervention details: Was fidelity absent or not reported?

Specification of key intervention details such as frequency, intensity, or duration is essential for both intervention fidelity and reproducibility of trials results.[133] The quality of reporting of key intervention details (CONSIDER checklist items 1-5) in this sample of complex intervention RCTs was poor, despite the introduction of extensions to CONSORT reporting guidelines such Standard

Protocol Items: Recommendations for Interventional Trials (SPIRIT) in 2013,[26] Template for Intervention Description and Replication (TIDieR)[150] in 2014, and the Consensus on Exercise Reporting Template (CERT)[151] in 2016. These guidelines were intended to improve the reporting of key aspects of intervention implementation in clinical trials and their protocols. Despite this, over 46 % of RCTs did not report at least one key aspect of the intervention materials, dosage, intensity, procedures, or the provider qualification or training received in the intervention, when assessed with the CONSIDER checklist items 1-5. Neary a third (31.9%) of studies failed to fully report (did not report or only partially reported) key aspects on intervention dosage (frequency, duration, intensity), and nearly one quarter (24%) of publications only partially described the intervention with enough detail to allow for replication.

This data supports previous findings that published clinical trial papers in complex interventions do not report interventions or trial procedures sufficiently to allow for reproducibility in practice or research.[222,388,486] Robinson, et al. found that details of surgical interventions were limited in more than the half (58.2%) of the RCTs examined, and not specified in 41 (10.6%).[90] Reporting of intervention details also remains poor for RCT publications in physiotherapy and rehabilitation as well, and systematic reviews of methodological quality in physiotherapy RCTs have typically excluded intervention description or implementation in their analyses.[222,487,488]

The poor reporting of intervention details also extends to reporting of intervention fidelity and fidelity monitoring. Although SPIRIT,[26] TIDieR,[150] and CERT[151] include items for participant adherence and intervention fidelity, only 67.2% of RCTs in this study report monitoring

participant adherence or any other aspect of intervention fidelity. Additionally, only 57 (27%) of publications in this study described modifications or tailoring of interventions taking place during the trial. These modifications were reported to be unplanned in nearly a third (18) of those. It is unknown in how many trials unplanned deviations from the protocol occurred but were not recorded or reported in trial publications. It is also unknown how many unpublished protocols include measures to monitor or maintain intervention fidelity that was not reported in trial publications.[157] Reported methods do not always reflect actual conduct within a trial.[489]

It is important to note that poor reporting quality in RCT publications may not reflect poor quality trial conduct, however.[490] Mhaskar et al., compared risk-of-bias assessments of 429 RCT protocols to assessments of their subsequent publications and found that the methods reported in the published manuscripts did not reflect the methods described in their protocols.[490] Poor reporting quality did not reflect the actual high methodological quality of eligible RCTs.[490] Researchers risk conflating the relationship between reporting of study conduct and treatment effect with that of actual conduct and treatment effect.

It is possible that intervention fidelity was monitored but not reported in RCTs finding robust treatment effects. Conversely, in trials resulting in small or equivocal intervention effect estimates, fidelity may have been reported to help explain the results. Indeed, we found that randomised controlled trials with moderate to large effect sizes reported monitoring fidelity less often than trials with small effect sizes (table 4.3). This can create a “chicken or egg” conundrum, in which it becomes difficult to determine if RCTs treatment effects influence fidelity reporting,

or if fidelity influences reported treatment effects. The reasons behind these differences in fidelity reporting cannot be inferred from our meta-analysis. These need exploration in an ongoing clinical trial, which would shed light on how fidelity actually works in practice and what may influence whether it is reported or not. This is explored in the following chapter (V), in which the CONSIDER framework and intervention fidelity are explored in a contemporary pragmatic randomised controlled trial comparing surgical and non-surgical management for the ACL deficient knee (ACL-SNNAP trial[349]).

## Strengths

This meta-epidemiological study addresses a gap in our understanding of the influence of trial characteristics on the outcomes of RCTs of complex interventions.[441,445,472] It quantified the influence of intervention fidelity and the quality of intervention delivery on treatment effect estimates in complex intervention RCTs. To the best of our knowledge, this is the first meta-epidemiological study to investigate this in complex interventions. Failure to examine intervention quality, in addition to methodological quality, has been identified as a limitation of systematic reviews of complex interventions.[472,483]

This study was rigorously developed following the recommendations of Sterne et al. in "Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological research[79]" and guidelines for meta-epidemiological research proposed by Murad and Wang.[450] It was also designed to avoid important pitfalls in meta-epidemiological

research and address the key recommendations made by Moustgaard et al. in "Ten questions to consider when interpreting results of a meta-epidemiological study—the MetaBLIND study.[80]" As a result, the search strategy was comprehensive, and full-text screening was undertaken to determine eligibility for every meta-analysis and RCT assessed for eligibility after screening by title and abstract. A power and sample size calculation developed specifically for meta-epidemiological studies was used to ensure adequate power.[436] That number was exceeded by 35% to overcome any unexpected discrepancies in estimates or elements of the sample size calculation, given that this is the first meta-epidemiological study of intervention fidelity, avoiding the inadequate sample size and low precision identified as frequent limitations of meta-epidemiological studies.[80] This is also one of the few meta-epidemiological studies examining continuous outcomes, addressing a gap in previous meta-epidemiological studies of interventions in healthcare.[80]

The CONSIDER checklist used to assess fidelity in this study was developed through extensive thematic and best-fit framework synthesis and its validity and reliability were established in previous stages of this program of research. The CONSIDER checklist allowed for multifaceted identification of fidelity in clinical trials despite the multiple descriptors used to describe fidelity in complex interventions literature. Trials' fidelity was also assessed blinded to trial effect estimates, reducing the potential for differential misclassification and over or underestimation of fidelity's influence on treatment effect estimate.

This study's large sample of 204 trials and 248 randomized comparisons also allowed for

subgroup analysis and meta-regression adjusted for key variables, specified *a priori*, while maintaining precision and robust estimates of fidelity's influence on treatment effect estimates. Meta-regression was limited to three key variables (sample size, risk of bias, and degree of fidelity), prevented over-fitting of our regression model.[491,492] Lack of adjustment for potential confounders is a limitation of the body of meta-epidemiological research.[80,82]

## Limitations

Despite this study's sample and power calculation and large sample size, it is possible that an even larger sample of meta-analyses and RCTs may alter the magnitude of fidelity's influence on treatment effect sizes. This study was conducted using only published data for completed trials. Given the potential risk of publication bias, our results may be conservative. If high levels of fidelity result in smaller effect estimates, it is possible that publication bias or outcome non-reporting may negatively impact the availability of data from those trials compared with trials in which fidelity was not monitored.[493] This may lead to reduced contrast[80] in effect estimates between the trials with and without fidelity monitoring. There is no feasible methodological approach to counter this risk in meta-epidemiological studies, however. Meta-epidemiological studies also layer three levels of analysis (participants, trials, and meta-analyses), and each of these has the potential to increase heterogeneity.

Reporting of intervention fidelity can often be limited or unclear, requiring a degree of subjective judgement by systematic reviewers and meta-epidemiological researchers. [45,169,290,387,388,494]

Overly narrow identification criteria may lead to a larger proportion of trials being classified as lacking fidelity, while overly broad criteria may dilute an effect, if present. The CONSIDER checklist used to assess fidelity in this study is comprehensive, utilizing multiple keywords and descriptors to identify fidelity processes in trial publications, allowing us to identify fidelity in RCT publications in multiple ways and limiting the potential for misclassification bias.

Additionally, the literature searches and data extraction were conducted by only one author (AP), and not undertaken in duplicate by other authors. This increases the risk that errors in search result eligibility screening or data extraction could be missed, or the author's biases may influence the results of the eligibility screening or data extraction. Future meta-epidemiological studies of intervention fidelity, or future development of the meta-epidemiological study in this thesis, should include multiple authors in the literature searches, eligibility screening, and data extraction to minimize these risks.

Finally, an inescapable limitation of this and other meta-epidemiological studies is that they assess the study characteristics that were reported in trial publications, rather than through direct observations of the trials themselves. Fidelity may have been monitored in a trial but not reported for several reasons. Editorial constraints, including word count limits and manuscript length restrictions, may also have limited reporting of intervention fidelity in the RCTs we assessed.[388,389] We attempted to overcome this limitation by also assessing RCT's protocols and trial registrations, when they were available, for additional information about fidelity monitoring

that was planned. Nevertheless, fidelity may have been monitored and maintained more frequently than our findings would suggest.

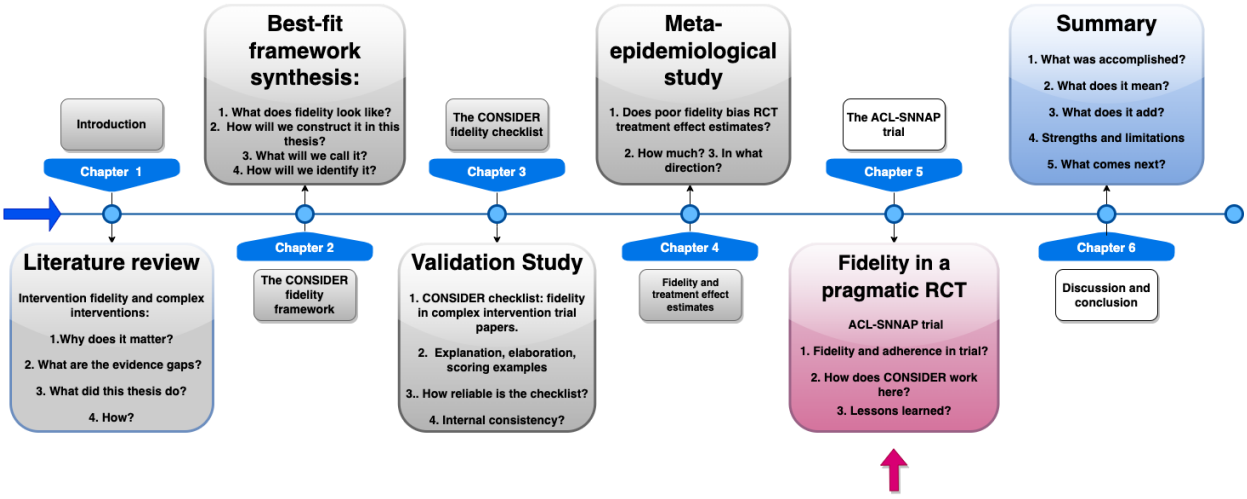
## CONCLUSION

This study quantifies, for the first time, the effect of poor intervention on treatment effect estimates derived from RCTs of complex interventions. The findings of this meta-epidemiological study suggest that lack of intervention fidelity in RCTs of complex interventions is associated with exaggerated and less precise observed effect estimates. Poor or absent intervention fidelity in RCTs may lead to overestimation of the magnitude of observed treatment effects, skewing the conclusions from individuals studies and systematic reviews with meta-analyses when pooled.[79,104] This could lead to the adoption of ineffective interventions, or premature rejection of interventions that could be effective. Our results show, empirically, that intervention fidelity is essential for accurate interpretation of intervention effects.[56]

Chapter V: The ACL-SNNAP trial: The CONSIDER framework and fidelity in a contemporary pragmatic randomised controlled trial.

## Chapter Summary:

The primary aim of this chapter was to gather insights about the CONSIDER framework and the results of the meta-epidemiological study through the greater level of granularity available from real-time data gathered prospectively in an ongoing clinical trial. In essence, we set out to "road test" the CONSIDER framework in the Anterior Cruciate Ligament (ACL) Surgery Necessity in Non-Acute Patients (ACL-SNNAP) trial. Key aspects of the framework and the meta-epidemiological study in Chapter IV were validated and several important insights about fidelity monitoring and fidelity reporting in complex intervention clinical trials were gained.



## BACKGROUND

Fidelity can act as a moderating factor between interventions and their intended outcomes.[24,37,50,51,495,496] The links between fidelity and intervention outcomes are complex, however. In the meta-epidemiological study of intervention fidelity in the previous chapter, poor interventions fidelity was associated with over-estimation of intervention effects and reduced precision. The direction of bias was unexpected, raising questions about factors influencing the relationship between intervention fidelity and treatment effect estimates derived from clinical trial publications.

The boundaries between poor intervention fidelity and poor treatment effectiveness can be unclear in rehabilitation clinical trials.[224,497,498] Like complex interventions, fidelity includes multiple, interacting components that can act synergistically or antagonistically to influence intervention effects.[40] Intervention fidelity constructs may also need to be flexible, allowing for intervention fidelity that accommodates (or can be tailored for) unique settings, contexts, and trial priorities.[216] Some components of intervention fidelity may be also be more essential in some kinds of trials or interventions than in others.[2,499,500] For example, monitoring tailoring of interventions to individual participants may be less pressing in tightly controlled explanatory trials than in pragmatic trials, which have more heterogenous study populations and a greater need for intervention tailoring.[210,225,498]

However, which fidelity components, or violations of fidelity, most influence intervention effects in clinical trials is uncertain. This can be of particular importance in pragmatic trials (pRCT), which

offer practitioners considerable leeway in adapting and applying interventions.[5,7,13,208] Key aspects of intervention delivery are less tightly controlled in pRCTs, creating challenges for monitoring and maintaining intervention fidelity.[210,216,224,225,498,501] Additionally, there is little consensus on the degree of fidelity that is needed, or appropriate, for pragmatic trials.[225,502,503]

Several other uncertainties about intervention fidelity in clinical trials remain. It is not certain how fidelity is reported in trial publications, or if its absence truly reflects trial actions and fidelity in real-world settings.[490] Less than two thirds of the RCTs assessed in the meta-epidemiological study in chapter IV reported monitoring fidelity, but it is unknown whether fidelity was truly absent or whether it was monitored but not reported in some of these trials. In trials with large treatment effects, fidelity may have been monitored but not reported, while in trials with small or equivocal effect, fidelity may have been monitored and reported to help explain the trial results.

These uncertainties are difficult to investigate through meta-epidemiological studies or secondary analysis of published data. Exploration of intervention fidelity in an ongoing clinical trial was needed to further understand how fidelity "works" in real-world, clinical trial settings and shed light on the findings of the meta-epidemiological study in chapter IV. This chapter describes both 1. our experience of using the CONSIDER framework to assess intervention fidelity in a contemporary pRCT (The Anterior Cruciate Ligament (ACL) Surgery Necessity in Non Acute Patients (ACL-SNNAP).[504] ) that was susceptible to complex fidelity issues by its nature on

behalf of the trial, and 2. provides insight into the "cause or effect" issue of effect size influence from fidelity adherence or reporting

This chapter explores the following AIMS:

AIM I. Apply the CONSIDER framework to assess fidelity and participant adherence in ACL-SNNAP (on behalf of the trial) to determine:

a. Can CONSIDER be used to evaluate the fidelity of a complex intervention pragmatic trial?

b. Describe the types of fidelity and adherence encountered in an ongoing clinical trial

AIM II: Describe how fidelity and participant adherence can affect the analysis and interpretation of ACL-SNNAP's results

AIM III. Glean insights about the results of the meta-epidemiological study through the experience of trial conduct and reporting in the ACL-SNNAP trial.

## Pragmatic randomised controlled trials (pRCTs) and intervention fidelity

Explanatory randomised trials are conducted under idealised conditions (for example double blinded, placebo-controlled) to give interventions the best chance to demonstrate an effect (efficacy).[390,391] By contrast, pragmatic randomised clinical trials are designed to evaluate the relative effectiveness of interventions under real-life conditions.[209] These can include diverse clinical populations and settings and comparison against usual care interventions.[210–212] While explanatory trials emphasise high levels of fidelity to a strict protocol, pRCTs allow for flexible

intervention delivery, varied participant adherence, and sometimes even choice of comparator, resembling clinical practice conditions.[210,225,498]

Clinical practice settings are complex by nature, however. Clinical populations are often more heterogenous than those in explanatory trials.[210,496,505,506] Changes to intervention delivery may also become necessary to accommodate individual participants or interventionists, real-world contexts or constraints, or unanticipated events. [210,496,505,506] Additionally, it is often unclear which intervention components, or their combinations, are modifiable and to what degree they can be adapted without compromising the intervention's ability to achieve the intended outcome.[499,500] The intervention components being modified may ultimately be central for achieving interventions' targeted outcomes.[499,500] These distinctions can be particularly challenging to make in complex interventions like physiotherapy, in which multiple interacting components and participants (patients and providers) can influence treatment delivery.[6,210] In pRCTs, those interventions are also often delivered by clinicians, rather than by trained researchers (as in many explanatory RCTs).[495,496] Variations in practitioner competence or intervention delivery can increase the scope for variations in intervention delivery.[495,496]

## Monitoring Fidelity in Complex Intervention Pragmatic Trials

Recommendations for monitoring and maintaining fidelity have mostly been focused on explanatory, rather than pragmatic trials. Little guidance exists for monitoring intervention fidelity and participant adherence in pragmatic trials.[507] There is no gold standard tool for assessing them.[507] Research into factors influencing intervention fidelity in pRCTs is also still incipient.[216,224,496,498,506] Much of the available research on fidelity in pragmatic trials has focused

on reporting the experiences of researchers in delivering interventions in pragmatic trials or making recommendations for fidelity when interventions are implemented in clinical practice, rather than a wider perspective and when investigated in a trial.[216,224,508,509] Frameworks such as Reach, Effectiveness, Adoption, Implementation, Maintenance (RE-AIM), the Framework for Reporting Adaptations and Modifications-Enhanced (FRAME)[47], and the Practical Robust Implementation and Sustainability Model (PRISM)[395] were developed to improve or support sustainable implementation of evidence-based interventions, but do not provide guidance for monitoring or assessing fidelity during clinical trials. The National Institutes of Health (NIH) Care Systems (HCS) Research Collaboratory workshop also explored challenges and recommended strategies for improving the dissemination, implementation, and sustainability of findings from pragmatic clinical trials, but these did not include strategies to enhance or monitoring intervention fidelity during pragmatic trials.[508]

What degree of fidelity is essential, or at which stage of a pRCT it should be monitored, (if at all) remain uncertain.[210] It is also not universally accepted that intervention fidelity should be enhanced, rather than monitored, or even monitored extensively in pRCTs.[225,502,503]Arguments against extensive fidelity monitoring in pragmatic trials include that strict standardization of interventions and provider characteristics does not reflect routine practice, reducing generalizability of the trial results to real-world settings.[225,502,503] Treatments in complex interventions such as surgery or physiotherapy are also multifaceted and it may not be clear which intervention components are important to measure or how they can be operationalized.[499,500] If fidelity or participant adherence are assessed in an ongoing manner, it

can also become obtrusive or time consuming, hindering clinicians' daily workflow and leading to poor participation by clinicians.[229,499,510]

From this perspective, it could be argued that there should be no, or only minimal, measurement of fidelity or participants adherence in pragmatic trials, as the act of monitoring adherence could change patients' treatment adherence or the behaviour of health professionals.[225,502] No additional strategies should be used to maintain or improve provider or participant adherence, unless also done in usual care.[478] Similarly, providing specialized training of intervention providers or monitoring their fidelity during the trial would not represent real-world practice, where similar levels of expertise or monitoring may not exist.[210,225]

Counter arguments to these positions emphasize that actions aiding in-trial fidelity, such treatment manuals, interventionist training, and monitoring of treatment delivery, can also be considered pragmatic (and helpful), as healthcare practitioners in clinical settings must also be competent in administering treatments to their patients and will often have at least some amount of training and supervision in clinical practice.[210,498] Additionally, routinely collected clinical documentation can also be used to monitor or assess intervention delivery without interrupting the intervention delivery or imposing undue burdens on treatment providers.[217,342] Pragmatic trials also lie along a continuum, from relatively controlled to more highly pragmatic.[391] Fidelity monitoring may be less obtrusive in more controlled than in highly pragmatic trials.

Arguments against fidelity monitoring in pRCTs may also miss the important contributions of intervention fidelity in understanding intervention effictiveness[152,226,271] Fidelity monitoring may be critical in multi-site pragmatic trials to assess if the same intervention is being studied across sites and settings. Monitoring fidelity and adherence in pragmatic trials is also essential for understanding barriers or enablers of patients' adherence to interventions in clinical practice.[511]

## Evidence gaps: Fidelity in pragmatic trials

Several important gaps in the evidence-base for intervention fidelity in pragmatic trials arise.[152,498] A paucity of research has investigated the challenges of assessing fidelity in these trials, particularly in complex interventions such as rehabilitation.[224,497] Previous studies have examined fidelity qualitatively, providing important information about factors influencing how fidelity were experienced in the trial, but limiting their usefulness for interpreting the effect of fidelity on quantitative outcomes.[152,216] There is also little guidance for monitoring intervention fidelity in rehabilitation pRCTs, and practical assessment of fidelity models applied in rehabilitation research is rare.[224,497,498] Researchers have also created intervention-specific fidelity tools for their studies without the guidance of an *a priori* theoretical framework.[224,497]

## The CONSIDER framework in a pRCT: The ACL-SNNAP Trial

The **Co**mplex I**n**terventions De**s**i**g**n, **D**elivery, **R**eceipt (CONSIDER) fidelity framework was developed as a multidimensional fidelity construct, recognizing the interactions between components of fidelity, such as Design (intervention and study), Delivery (intervention delivery fidelity), and Receipt (participant adherence) (figure 5.1). The applicability of the framework to

ongoing clinical trials was investigated in the Anterior Cruciate Ligament (ACL) Surgery Necessity in Non-Acute Patients (ACL-SNNAP) trial on behalf of, and as part of, the trial.[349] The ACL-SNNAP trial provided a unique opportunity to study issues of fidelity in depth in a pragmatic rehabilitation trial.

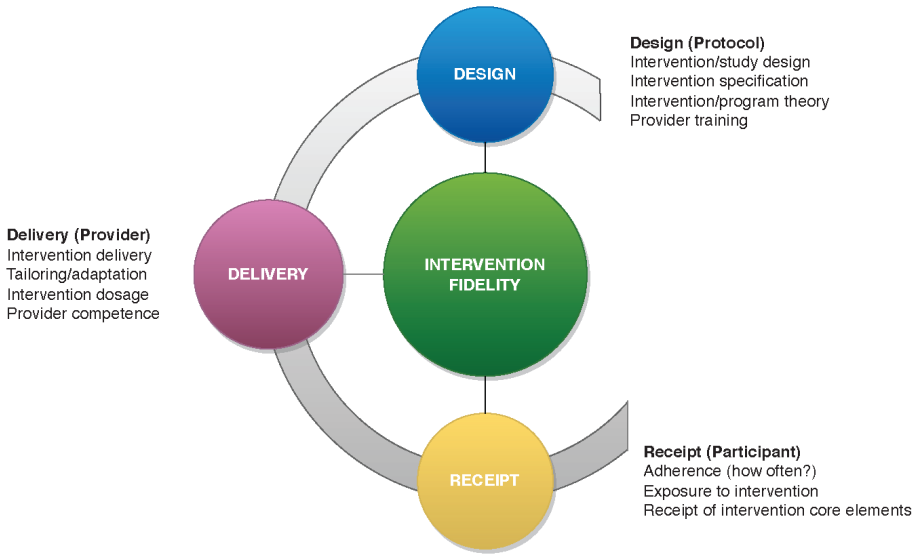



**Figure 5.1: Co**mplex I**n**terventions De**si**gn, **De**livery, **R**eceipt (CONSIDER) fidelity framework

## ACL-SNNAP

The ACL-SNNAP trial is a pragmatic, randomised controlled trial comparing the clinical and cost effectiveness of two management strategies for non-acute Anterior Cruciate Ligament (ACL) injury: Rehabilitation versus surgical Reconstruction.[349] Up to 300,000 persons per year experience a ruptured ACL ligament in the United States, with an annual incidence of 68.6 per 100,000 person-years.[512–514] ACL injuries can lead to prolonged or recurrent knee instability, reduced physical activity and quality of life, and increased risk of knee osteoarthritis.[515,516] Some patients with ACL tears recover from their initial injury and return to daily function without an

intact ACL after physiotherapy. Others continue to experience knee instability and are referred for surgical ACL reconstruction to treat functional knee instability.[516]

Although surgical management is the preferred treatment for ACL injuries in the UK, with 15,000 to 50,000 ACL reconstruction surgeries performed yearly in England,[514] current management of ACL injury is based on limited evidence.[504,517,518] A systematic review by Monk et al. in 2016 found insufficient evidence from randomised trials to determine whether surgery or non-surgical management was more effective for ACL injury.[504] Low-quality evidence suggested no difference between surgical management (ACL reconstruction followed by structured rehabilitation) and non-surgical treatment (structured rehabilitation only) in patient-reported outcomes and knee function at two and five years after injury.[504] More recent evidence also questioned the preference for surgical management of ACL rupture. Frobel et al., investigated five-year patient reported and radiographic outcomes between 121 young, active adults with acute ACL tear treated with either rehabilitation plus early ACL reconstruction and those treated with rehabilitation and optional delayed ACL reconstruction.[519] Rehabilitation plus early ACL reconstruction did not provide better results at five years than initial rehabilitation with the option of later ACL reconstruction. The most effective management strategy for chronic ACL injury remains uncertain.[504,520]

## ACL-SNNAP overview

The primary objective of the ACL-SNNAP study was to determine whether a strategy of non-

surgical management [Rehabilitation] (with option for later ACL reconstruction only if required) is more clinically effective and cost effective than a strategy of surgical management (reconstruction) in patients with non-acute (more longstanding) Anterior Cruciate Ligament Deficiency (ACLD).[504]

Trial design: The ACL-SNNAP study was a pragmatic multi-centre randomised controlled trial with two-arm parallel groups and 1:1 allocation ratio.[349] On the Pragmatic Explanatory Continuum Indicator Summary (PRECIS-2)[391], version 2 (table 5.1) ACL-SNNAP has an average score of 4.4 across the indicator's 9 items, representing a highly pragmatic trial. It compared non-surgical management (physiotherapy) and surgical management (reconstruction) options for persons with a symptomatic, non-acute ACL deficient knee.[349] Neither participants nor health care practitioners (surgeons and physiotherapists) were blinded.

| Domain | Comment |
|---|---|
| 1. Eligibility | Who is selected to participate in the trial? |
| 2. Recruitment | How are participants recruited into the trial? |
| 3. Setting | Where is the trial being done? |
| 4. Organisation | What experience and resources are needed to deliver the intervention? |
| 5. Flexibility: delivery | How should the intervention be delivered? |
| 6. Flexibility: adherence | What measure are in place to make sure participants adhere to intervention? |
| 7. Follow-up | How closely participants are followed-up? |
| 8. Primary outcome | How relevant is to participants? |
| 9. Primary analysis | To what extent are all data included? |
| **Score (**each domain**):** from 1 to 5 using a 5-point Likert | 1 = very explanatory, 3 = equally pragmatic and explanatory, 5 = very pragmatic |

Table 5.1: The PRECIS-2 Framework[391]

Population: 316 trial participants with a symptomatic non-acute ACL deficient knee were recruited from 30 NHS orthopaedic units across the UK between 1st February 2017 and 12th April

2020. Of these, 156 participants were randomised to the Surgical management arm and 160 to the Rehabilitation management arm (figure 5.2). Eligible participants were at least 18 years old, with symptomatic ACL deficiency confirmed by clinical assessment and MRI scan.[504] Detailed eligibility criteria and study procedures are found in the ACL-SNNAP study protocol.[504]

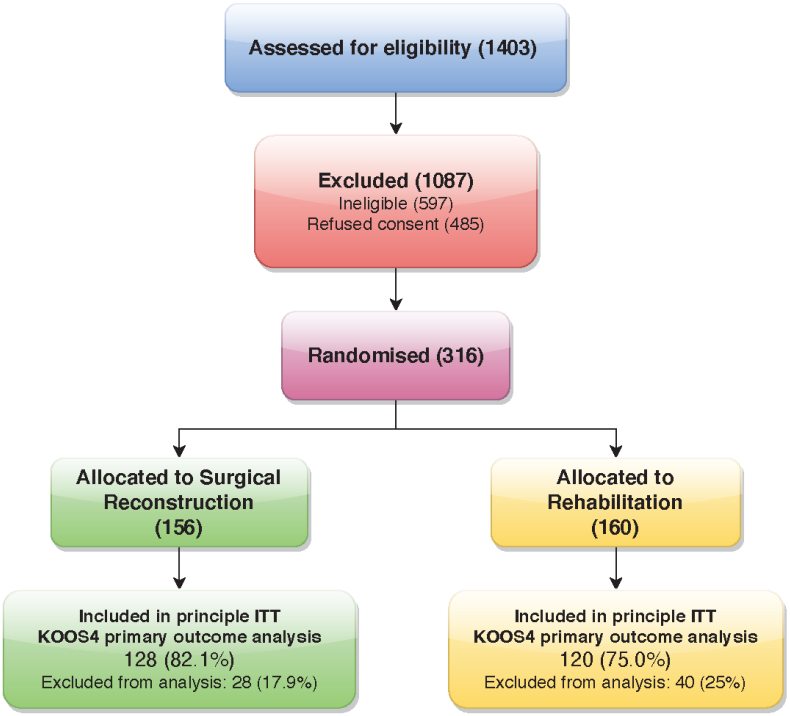



**Figure 5.2:** ACL-SNNAP flow chart. Adapted from Beard et al., 2022[349]

## Intervention and Comparison:

The interventions in both arms were routine NHS treatments. The description and standardisation of the trials' interventions were developed by review of evidence, results of a survey of ACL surgeons, synthesis of current practice guidelines, rehabilitation protocols from UK Trusts, and consensus meetings with surgeons and physiotherapists.[504] The content of both arms was based on a minimal set of pre-established criteria, allowing for variations in clinical practice

between both surgeons and physiotherapists while preserving the integrity of comparisons between arms.[504] In the non-surgical management (Rehabilitation) arm attended physiotherapy and were only listed for reconstructive surgery if they experienced continued knee instability or symptoms following rehabilitation. This is described in greater detail later in this chapter. The routine rehabilitation protocol used at participating sites was followed, with the addition of the following mandatory, trial-level aims:

1. Provision of at least 6 rehabilitation sessions over at least a 3-month period.
2. The rehabilitation program had to include the following components:
    a. Control of pain and swelling.
    b. Regaining range of movement.
    c. Improving neuromuscular control.
    d. Regaining muscle strength.
    e. Achieving normal gait pattern.
    f. Returning to function/activity/sport.
    g. Return to sport criteria.
3. Clearly identified progression milestones.
4. Identification criteria for poor or non-progression.

Rehabilitation was provided, or closely supervised, by senior physiotherapists experienced with ACL injury (table 5.5). A physiotherapy case report form (CRF) was used to document rehabilitation interventions and individuals' progression (appendix V figures 5.1,5.2). As there is little consensus for the most effective rehabilitation protocol for ACL injuries,[504] variation in the

specific exercises and adjuncts modalities used, and flexible adaptation of treatments to individuals were allowed. There were no pre-specified timelines for progression of therapeutic activities.

Participants randomised to the surgical management arm were referred for surgical ACL reconstruction (using any technique chosen by the surgeon) without any further rehabilitation. Participants in the surgical arm could receive pre-surgical physiotherapy to address acute symptoms, such as swelling, but not a formal ACL rehabilitation programme beyond basic maintenance exercises before surgery. Surgical reconstruction could utilize either patellar or hamstring tendon graft, depending on surgeons' preference. Detailed information about the surgical procedures can be found in the study protocol.[504] Surgeons performing the surgical reconstruction had expertise in soft tissue knee surgery as indicated in the Best Practice for Primary Isolated Anterior Cruciate Ligament Reconstruction guidelines[521], with a minimum experience of 50 previous procedures. Participants undergoing surgical reconstruction were referred for a programme of post-operative rehabilitation (physiotherapy) as per standard care at the participating hospital. The initial content of post-operative physiotherapy differed from non-surgical management physiotherapy by including aspects of graft protection and caution necessary following ACL reconstruction.[504]

Outcomes: The primary outcome was participants' Knee injury and Osteoarthritis Outcome Score (KOOS) at 18 months post-randomisation. Secondary outcomes included return to sport or

activity, intervention related complications, patient satisfaction, expectations of activity, generic health quality of life, knee specific quality of life and resource usage.[349]

Statistical Analyses: A detailed statistical analysis plan was published for the trial.[522] Briefly, all principal outcome analyses were based on the intention-to-treat (ITT) principle. Participant data was analysed in the groups to which they were randomly assigned, regardless of their adherence with treatment allocation. Participants in the rehabilitation arm who required surgery were analysed as randomised, as this was an expected outcome and was part of the protocol. In the surgical reconstruction arm, "Reconstruction or Surgery" referred to a decision to list for surgical reconstruction and not necessarily the point in time of the surgical procedure.

It was also anticipated that the complex nature of both intervention arms would lead to several potential treatment pathways.[522] For example, it was expected that some participants would not complete their allocated treatment, whether by not starting rehabilitation, having insufficient rehabilitation, not having reconstructive surgery within 12 months in the surgical management group, not having sufficient post-operative rehabilitation (at least 5 months of formal identifiable post op therapy), or not having reconstructive surgery but crossing over to rehabilitation instead (figure 5.3). Additionally, some participants may undergo delayed surgery, leaving insufficient follow up time or post-operative rehabilitation during the trial. Some participants in the rehabilitation arm may also elect or require surgery, while others may not complete their allocated or intended treatment. Participant treatment profiles were constructed, *a priori*, to capture these expected deviations from their allocated treatment pathway (figure 5.3).[522]

Comprehensive **per protocol analyses** had been constructed, *a priori*, as part of the trial's statistical analysis plan[522] to enable investigation of the impact of participant adherence on study outcomes and fair comparisons and interpretations in the event of poor adherence. Two primary per protocol analyses were conducted.[522] Both excluded participants (in both arms) who did not fulfil minimal protocol criteria. A **conservative per protocol ($PP_C$)** analysis excluded participants who did not meet the minimum requirements for the trial in their intervention arm (table 5.7).[522] The PPc represents a focus on study or experimental design, considering protocol non-adherence. A second, **pragmatic per protocol (PPp)** analysis included participants who started but underwent insufficient physiotherapy or did not complete physiotherapy treatment, even if they went on to have surgery.[522] These participants had poor adherence to rehabilitation. Their inclusion in the PPp represents pragmatic features of the trial and real-world care, reflecting patterns of participant adherence to treatment typically seen in the NHS.

**Setting:** Participating NHS sites were selected based on having an established practice of ACL reconstruction and an experienced ACL reconstruction knee surgeon and physiotherapy team capable of providing contemporary care.

**Trial Results:**

The ACL-SNNAP trial found that surgical reconstruction is a clinically superior and more cost-effective management strategy for persons with non-acute ACL injury and persistent knee instability compared to rehabilitation management (table 5.7).[349] Intention to treat (ITT) analyses resulted in mean KOOS4 scores at 18 months of 73.0 (SD 18.3) in the surgical group, compared

to 64.6 (21.6) in the rehabilitation group, with an adjusted mean difference of 7.9 (95% CI 2.5, 13.2; p=0·0053) in favour of surgical management.[349] The unadjusted analyses also produced statistically significant effects in favour of surgical management, resulting in an adjusted mean difference of 8.3 (3·3, 13.3), p=0·0012.[349] Both the PPp and PPc analyses also supported the ITT results, with all treatment effects favouring surgical management at a level reaching statistical significance.[349] Of note, both treatment groups did improve over time, despite the superiority of surgical management over rehabilitation.

## AIM I: Apply CONSIDER to assess fidelity and participant adherence in ACL-SNNAP

CONSIDER (figure 5.1), was used to assess fidelity and adherence during the trial on behalf of, and as part of the trial.[523] Intervention delivery fidelity and intervention dosage (CONSIDER Delivery), and participant adherence (CONSIDER Receipt) were assessed with data from the 1677 physiotherapy and surgical case report forms (appendix V figures 5.1, 5.2) collected for 225 participants across 29 participating sites in the UK. These recorded participant adherence with their rehabilitation program and the fidelity of their rehabilitation programs to the mandatory aims and goals of the rehabilitation arm. However, in cases where there was missing data and/or to confirm whether the participant had undergone any treatment, the investigators contacted the participating site (hospital) or the treating physiotherapist to gather data about the physiotherapy treatment , the number of sessions and dates when the participant attended physiotherapy.

Anonymized data for participants' treatment session attendance (physiotherapy), treatment site,

session duration, session content as categorized as in the CRF, physiotherapist staff grade, discharge date and reason for discharge, and therapist comments was extracted into a Microsoft Excel spreadsheet for analysis. Descriptive statistics were calculated in Microsoft Excel for the number of CRFs collected, number of rehabilitation treatment sessions attended, dates of attendance, length (dose), content (type of treatment), and reason for non-attendance or discharge from physiotherapy treatment sessions.

Descriptions of therapeutic interventions as well as anonymized descriptive information about adherence, participant progression, or reasons for discharge from rehabilitation were extracted from the CRFs and imported into the NVIVO12 software package. Thematic analysis was used to code data extracted from the CRFs.[248] Thematic analysis followed the guidance and recommended steps of Saldana's (2016) Coding Manual for Qualitative Researchers, and the Streamlined Codes-to Theory Model for qualitative inquiry.[245] First-cycle coding codes for participant adherence and intervention were formulated through a review of the CRF forms using a descriptive (topic) coding method.[246] In second-cycle coding, pattern coding was used to reanalyse and reconfigure the first-cycle codes categorically and conceptually into a smaller number of pattern codes (themes).[247] Codes were then grouped into themes for factors influencing participant adherence. Each theme was then supported with corresponding excerpts from the CRFs .

### Intervention fidelity: Fidelity of delivery

Intervention fidelity assessment was not straightforward in ACL-SNNAP. The trial was highly

pragmatic and compared participant outcomes for two management strategies, rather than two specific treatments. A flexible, pragmatic approach to the assessment of intervention fidelity and adherence was needed that reflected the goals and context of the trial.

### Complex fidelity in ACL-SNNAP

As a movement-impairment management strategy, physiotherapy (the rehabilitation arm) includes a vast array of interventions and a number therapeutic mechanisms which may be poorly defined.[332] Given this and the need to generalize ACL-SNNAP's findings to a broad range of persons, settings, and treatment methods, interventions within the rehabilitation arm were not strictly defined. This allowed for a wide range of physiotherapy interventions targeting rehabilitation of the ACL-deficient knee and progression criteria, reflecting the diversity seen in routine clinical practice.[524]

Over sixty therapeutic interventions were listed as notes or addenda in the physiotherapy CRFs. These ranged from acupressure and wound-care to soft-tissue neural mobilization, blood flow restriction therapy, and plyometric exercise, covering a wide range of potential active ingredients and therapeutic mechanisms. Additionally, the rehabilitation CRFs included intervention categories, based on the desired progression for the trial (tables 5.2: representing the activities specified in the study protocol, and 5.3: what was actually delivered), but interventionists did not routinely report individual therapeutic interventions performed during treatment sessions.

| | Stages of Rehabilitation | | |
|---|---|---|---|
| Mandatory aims/goals of the rehab. intervention | 0-6 weeks | 7-12 weeks | >3 months |
| 1. Control of pain and swelling | Interventions applied to treat pain and swelling, as required.<br><br>Hydrotherapy<br>Advice and education<br>Other | Education on activity modification to address activity related swelling, if required.<br>n/a<br>Advice and education<br>Other | Education on activity modification to address activity related swelling, if required<br>n/a<br>Advice and education<br>Other |
| 2. Regain range of movement (ROM) | Progressive exercises to facilitate ROM.<br>Supervised exercise:: stretching<br><br>Home exercise (instruction/review)<br>Other | <br>n/a<br><br>n/a<br><br>Other | <br>n/a<br><br>n/a<br><br>Other |
| 3. Improve neuromuscular control | Exercises aimed at improving static balance.<br><br>Supervised exercise: proprioception<br><br>Supervised exercise: strength<br>Home exercise (instruction/review)<br>Other | Exercises progressed to multi angle static balance and dynamic balance.<br>Supervised exercise: proprioception<br><br>Supervised exercise:: strength<br>Home exercise (instruction/review)<br>Other | Further progression of dynamic balance exercises and sport specific activities.<br>Supervised exercise: proprioception<br><br>Supervised exercise:: strength<br>Home exercise (instruction/review)<br>Other |
| 4. Regain muscle strength | Bilateral weight bearing activities. Exercises to improve quadriceps strength: unloaded full activation. Exercises to improve hamstrings strength: loaded exercise. Exercises to improve strength of other muscles of lower limb and trunk.<br><br>Supervised exercise:: strength<br>Home exercise (instruction/review)<br>Other | Progression of exercises to improve quadriceps strength: closed chain exercises without limitations. Progression of exercises to improve hamstrings strength: exercises without limitations. Progression of exercises to improve strength of other muscles of lower limb/trunk.<br><br>Supervised exercise: strength<br>Home exercise<br><br>Other | Progression of exercises to improve quadriceps strength: open chain exercises without limitations. Progression of exercises to improve hamstrings strength: exercises without limitations. Progression of exercises to improve strength of other muscles of lower limb trunk.<br><br>Supervised exercise:: strength<br>Home exercise (instruction/review)<br>Other |
| 5. Achieve normal gait pattern | Education to achieve normal gait pattern. | On achieving normal symmetrical gait pattern, | Progression to running on flat/even surface. |

| | | | |
|---|---|---|---|
| | Gait re-education<br><br>Other | progression to fast walking/jogging (flat/even surface).<br><br>Supervised exercise:: strength<br>Other | Progressed to unrestricted running.<br><br>Supervised exercise:: strength<br>Other |
| 6. Return to function/ activity/sport | Focus on regaining knee function and achieving normal symmetrical gait pattern. No sport specific activities. Static cycling without restrictions<br><br>Supervised exercise:: strength<br><br>Home exercise (instruction/review)<br>Other<br>Individual | Progression to controlled/ predictable sports and recreational activities.<br>Progression to outdoor cycling without re-strictions, cross trainer, swimming (front crawl).<br>Supervised exercise:: strength<br><br>Supervised exercise:: sport specific<br>Home exercise<br><br>Other<br>Individual/Group | Gradual introduction to sport specific exercises. Sport specific running agility running drills, sport specific skills.<br><br>Supervised exercise:: strength<br><br>Supervised exercise:: sport specific<br>Home exercise (instruction/review)<br>Other<br>Individual/Group |

Table 5.2: Rehabilitation content and progression of activities, rehabilitation

| Content of rehabilitation sessions (pre-surgery) | Surgical Management n (%) | Rehabilitation n (%) | Total n (%) |
|---|---|---|---|
| Total number participants | 14 | 121 | 135 |
| Total number of sessions | 64 | 751 sessions | 801 sessions |
| Advice and education | 63 (98.4) | 672 (89.5) | 735 (90.2) |
| Supervised exercises (strengthening) | 58 (90.6) | 663 (88.3) | 721 (88.5) |
| Supervised exercises (stretching) | 38 (59.4) | 377 (50.2) | 415 (50.9) |
| Supervised exercises (sport specific) | 8 (12.5) | 240 (32.0) | 248 (30.4) |
| Home exercises (instructions/review) | 60 (93.8) | 471 (62.7) | 531 (65.2) |
| Gait re-education | 10 (15.6) | 108 (14.4) | 118 (14.5) |
| Supervised exercises (proprioception) | 48 (75.0) | 565 (75.2) | 613 (75.2) |
| Hydrotherapy | 0 (0.0) | 9 (1.2) | 9 (1.1) |
| **Content of rehabilitation sessions (postsurgery)** | **Surgical Management n (%)** | **Rehabilitation n (%)** | **Total n (%)** |
| Total number participants | 58 | 28 | 86 |
| Total number of sessions | 562 | 239 | 801 |
| Advice and education | 476 (84.7) | 204 (85.4) | 680 (84.9) |
| Supervised exercises (strengthening) | 472 (84.0) | 167 (69.9) | 639 (79.8) |
| Supervised exercises (stretching) | 351 (62.5) | 119 (49.8) | 470 (58.7) |
| Supervised exercises (sport specific) | 129 (23.0) | 102 (42.7) | 231 (28.8) |
| Home exercises (instructions/review) | 397 (70.6) | 146 (61.1) | 543 (67.8) |
| Gait re-education | 142 (25.3) | 84 (35.1) | 226 (28.2) |
| Supervised exercises (proprioception) | 302 (53.7) | 111 (46.4) | 413 (51.6) |
| Hydrotherapy | 7 (1.2) | 2 (0.8) | 9 (1.1) |

**Table 5.3:** Intervention category totals in pooled CRFs for pre- and post-surgical rehabilitation, both arms.

Analysing rehabilitation and surgical arm CRFs, the frequency of interventions delivered or performed in each category were tabulated across participants (table 5.3). However, this did not provide sufficient information to assess intervention fidelity. Given the wide range of interventions, active ingredients, and dosages (described later in this section), it was not possible

to evaluate fidelity at the level of individual participant sessions or interventions. Another way to assess intervention fidelity in ACL-SNNAP was needed.

### AIM Ia: Can CONSIDER be used to evaluate fidelity in a complex intervention pragmatic trial?

Using CONSIDER, it was possible to find alternate ways to measure fidelity by considering the trial's goals and context. Fidelity was assessed at a program level as having fidelity to the rehabilitation arm's program theory (part of CONSIDER Design and Delivery) rather than at the level of individual interventions. A program theory provides a model for how interventions should be structured or administered to achieve a therapeutic outcome and goals that define the structure, process, and outcomes of a program.[339]

While the ACL-SNNAP trial protocol did not specify which interventions could be delivered in the rehabilitation arm, it did specify a desired progression of therapeutic activities toward those targeting return to sports or functional activities (table 5.2), representing the rehabilitation arm's program theory.[504] Fidelity in ACL-SNNAP was assessed by investigating the number of participants whose rehabilitation program included a progression towards supervised, sports specific exercises on the case report forms, or fidelity to ACL-SNAAP's program theory. Analysis of rehabilitation CRFs found interventions reported in the return to sport category included cycling, impact-treadmill running, progressive football-related exercise, on-pitch exercises, agility training (sport), sports-training exercises, and swimming, amongst others.

Fidelity was coded categorically:

1. **Fidelity not maintained**: Rehabilitation program did not include return to sports or functional activity (lacked fidelity to the rehabilitation arm's program theory).

2. **Fidelity maintained**: Participates completed the minimum adherence (60 days+) and had a rehabilitation program including a progression toward return to sports and function.

3. **High degree of fidelity**: Participants completed the minimum adherence (2 or more sessions of physiotherapy over more than a 60-day period) and a rehabilitation program that included a progression toward return to sport or functional activity.

### AIM Ib Describe the types of fidelity and adherence experienced in an ongoing clinical trial

The resulting analyses found that Intervention fidelity was moderately high in the rehabilitation arm (table 5.4).[138] Seventy-two percent of participants who were randomized to, and completed, any amount of physiotherapy met fidelity criteria (table 5.4). Additionally, the physiotherapy programs of 55 of the 95 (58%) participants with high adherence also included a progression to sport-specific exercises (table 5.4), for a moderate degree of fidelity.

| Intervention arm | Group | n | Total sessions | Mean (SD) n sessions | Range | Compliance or fidelity rate (%) |
|---|---|---|---|---|---|---|
| **Non-surgical arm** | Participants randomised to rehabilitation arm | 160 | 1039 | 6.3 (6.7) median 4 | 0-35 | |
| | Participants completed any amount of rehab | 125 | 1039 | 8.0 (6.7) median 6 | 2-35 | **78.0 %** |
| **Adherence** Physiotherapy takes place over ≥ 60-day timespan | | | | | | |
| **All adherence** | Participants were randomised to the nonsurgical arm (rehab), completed any amount of physiotherapy (> 2 sessions, >60 days) | 125/160 | 1039 | 8.0 (6.7) median 6 | 2-35 | **78.1 %** |
| **Low/partial adherence** | Participants were randomized to and started rehabilitation but did not meet adherence criteria | 34/160 | - | - | - | **21.3 %** |
| **High adherence** | Participants randomised to physiotherapy, met compliance criteria. | 95/160 | 959 | 10.1 (7.6) median 8 | 3-35 | **59.4 %** |
| | Among participants randomized to physiotherapy who began it, met adherence criteria | 95/125 | 959 | 10.1 (7.6) median 8 | 3-35 | **76.0 %** |
| **Fidelity** Physiotherapy included progression of activities toward return to sport or functional activity | Participants who completed any amount of rehab and rehab. Program met fidelity criteria. | 90/125 | - | - | 3-35 | **72.0 %** |
| **High degree of fidelity** | Participants had high compliance to rehab and physiotherapy program met fidelity criteria. | 55/95 | 594 | 11.6 (8.5) median 10 | 2-35 | **58.0 %** |
| **Number of days in rehab** | Pts who completed any amount of physiotherapy | 125 | 16825 | 136.8 (111.4) median 113 | 1-476 | |
| | Participants with high compliance (≥ 60-day timespan) | 95 | 16172 | 170.2 (104.6) median 136 | 60-476 | |
| **Surgical arm** | Participants randomized to the surgical reconstruction arm | 156 | | | | |
| **Adherence** | Participants randomized to and had surgical reconstruction | 113/156 | 635 | 10.08 | 1-31 | **72%** |
| **High adherence** | Participants had surgical reconst. And completed post-surgical rehab. Over at least a 5-month timespan. | 110/156 | 621 | 9.3 (7.17) | 1-31 | **70.5%** |
| **Low/partial adherence** | Had surgical reconstruction, but insufficient or no rehab | 3/156 | 14 | 1.55 (.53) | 0-2 | 2% |

Table 5.4: Fidelity and adherence data

Intervention dosage (CONSIDER Delivery):

Determining the Intervention dosage received by participants presented a particular challenge in fidelity assessment in ACL-SNNAP (table 5.5). Amongst the 121 participants in the rehabilitation arm receiving physiotherapy, the average number of individual or group physiotherapy sessions received was 5, but the range of sessions was wide at 2-35. The most frequent number of sessions attended was 6. Once group exercise classes were removed, the average duration of an individual session was 41 minutes, with a range of 5-60 minutes. The most frequently recorded sessions length (mode) was 30 minutes.

| REHABILIATION ARM | Surgical Management | Non-Surgical (Rehabilitation) | Total |
|---|---|---|---|
| **Received rehabilitation n (%)** | 14/156 (9.0) | 121/160 (75.6) | 135/316 (42.7) |
| **Average number of sessions prior to surgery (if had), (median IQR)** | 3 (2, 5) | 5 (3, 12) | 4 (3, 7) |
| **Total number of sessions pre-surgery (all participants)** | 64 | 751 | 815 |
| **Staff grade, n (%)** | | | |
| 5 | 10 (15.6) | 81 (10.8) | 91 (11.2) |
| 6 | 38 (59.4) | 376 (50.1) | 414 (50.8) |
| 7 | 15 (23.4) | 166 (22.1) | 181 (22.2) |
| 8 | 1 (1.6) | 61 (8.1) | 62 (7.6) |
| Other | 0 (0.0) | 67 (8.9) | 67 (8.2) |
| **Session length, mean (SD)** | 45.0 (13.2) | 40.2 (13.1) | 40.7 (13.2) |
| **Total session time per person mean (SD)** | 212.5 (297.4) | 259.5 (345.6) | 254.6 (340.2) |
| **Type of session, n (%)** | | | |
| One-to-one | 31 (48.4) | 436 (58.1) | 467 (57.3) |
| Group based | 33 (51.6) | 315 (41.9) | 348 (42.7) |
| **Session content, n (% of total sessions)** | | | |
| Advice and education | 63 (98.4) | 672 (89.5) | 735 (90.2) |
| Supervised exercises (strengthening) | 58 (90.6) | 663 (88.3) | 721 (88.5) |
| Supervised exercises (stretching) | 38 (59.4) | 377 (50.2) | 415 (50.9) |
| Supervised exercises (sport specific) | 8 (12.5) | 240 (32.0) | 248 (30.4) |
| Home exercises (instructions/review) | 60 (93.8) | 471 (62.7) | 531 (65.2) |
| Gait re-education | 10 (15.6) | 108 (14.4) | 118 (14.5) |
| Supervised exercises (proprioception) | 48 (75.0) | 565 (75.2) | 613 (75.2) |
| Hydrotherapy | 0 (0.0) | 9 (1.2) | 9 (1.1) |
| POST-SURGERY REHAB | Surgical Management | Non-Surgical (Rehabilitation) | Total |
| **Received rehabilitation n (%)** | 58/105 (55.1) | 28/52 (53.8) | 86/316 (27.2) |
| **Median number of sessions prior to surgery (IQ range)** | 8 (4, 16) | 7 (2.5, 12) | 8 (4, 13) |
| **Total number of sessions pre-surgery (all participants)** | 562 | 239 | 801 |
| **Staff grade, n (%)** | | | |
| 5 | 52 (9.3) | 17 (7.1) | 69 (8.6) |
| 6 | 232 (41.3) | 131 (54.8) | 363 (45.3) |
| 7 | 141 (25.1) | 51 (21.3) | 192 (24.0) |
| 8 | 53 (9.4) | 10 (4.2) | 63 (7.9) |
| Other | 69 (12.3) | 6 (2.5) | 75 (9.4) |
| **Session length, mean (SD)** | 40.6 (14.1) | 41.7 (15.6) | 41.0 (14.5) |
| **Total session time per person mean (SD)** | 357.8 (312.8) | 332.0 (402.3) | 349.4 (342.3) |
| **Type of session, n (%)** | | | |
| One-to-one | 375 (66.7) | 144 (60.3) | 519 (64.8) |
| Group based | 172 (30.6) | 71 (29.7) | 243 (30.3) |
| **Session content, n (% of total sessions)** | | | |
| Advice and education | 476 (84.7) | 204 (85.4) | 680 (84.9) |
| Supervised exercises (strengthening) | 472 (84.0) | 167 (69.9) | 639 (79.8) |
| Supervised exercises (stretching) | 351 (62.5) | 119 (49.8) | 470 (58.7) |
| Supervised exercises (sport specific) | 129 (23.0) | 102 (42.7) | 231 (28.8) |
| Home exercises (instructions/review) | 397 (70.6) | 146 (61.1) | 543 (67.8) |
| Gait re-education | 142 (25.3) | 84 (35.1) | 226 (28.2) |
| Supervised exercises (proprioception) | 302 (53.7) | 111 (46.4) | 413 (51.6) |
| Hydrotherapy | 7 (1.2) | 2 (0.8) | 9 (1.1) |

**Table 5.5:** Participant adherence and intervention program characteristics

## Participant adherence

Following the CONSIDER framework, participant adherence was defined as participants' initiation of their allocated intervention strategy and meeting the prescribed level of attendance in intervention sessions or trial activities. In the ACL-SNNAP trial, this describes whether participants began physiotherapy and participated in therapy sessions at the frequency required by the study protocol (at least 2 sessions over a 60-day period, with at least 6 total treatment sessions over a 90-day period).[504] Adherence was classified as:

### Non-surgical (rehabilitation) arm:

1. All: Participants allocated to and completed any amount of rehabilitation.
2. High: Participants allocated to and completed rehabilitation per the mandatory aims of the protocol, with at least 2 sessions over at least a 60-day period.
3. Partial (any): Participants allocated to and completed rehabilitation but not over a 60-day period. This included participants who started rehabilitation too close to final 18-month study completion date to have therapy over 60 or more days.
4. Insufficient: Participants attended less than 2 therapy sessions (insufficient rehabilitation).
5. None: Participants were allocated to rehabilitation but did not start it (no record of physiotherapy attendance).

### Surgical arm:

1. Participants allocated to surgery and had surgery and post -surgical rehabilitation

over at least a 5-month timespan.

2. Partial: Participants allocated to surgery and had surgery and insufficient or no post-operative rehabilitation.
3. None: Participants were allocated to surgery but never had it.

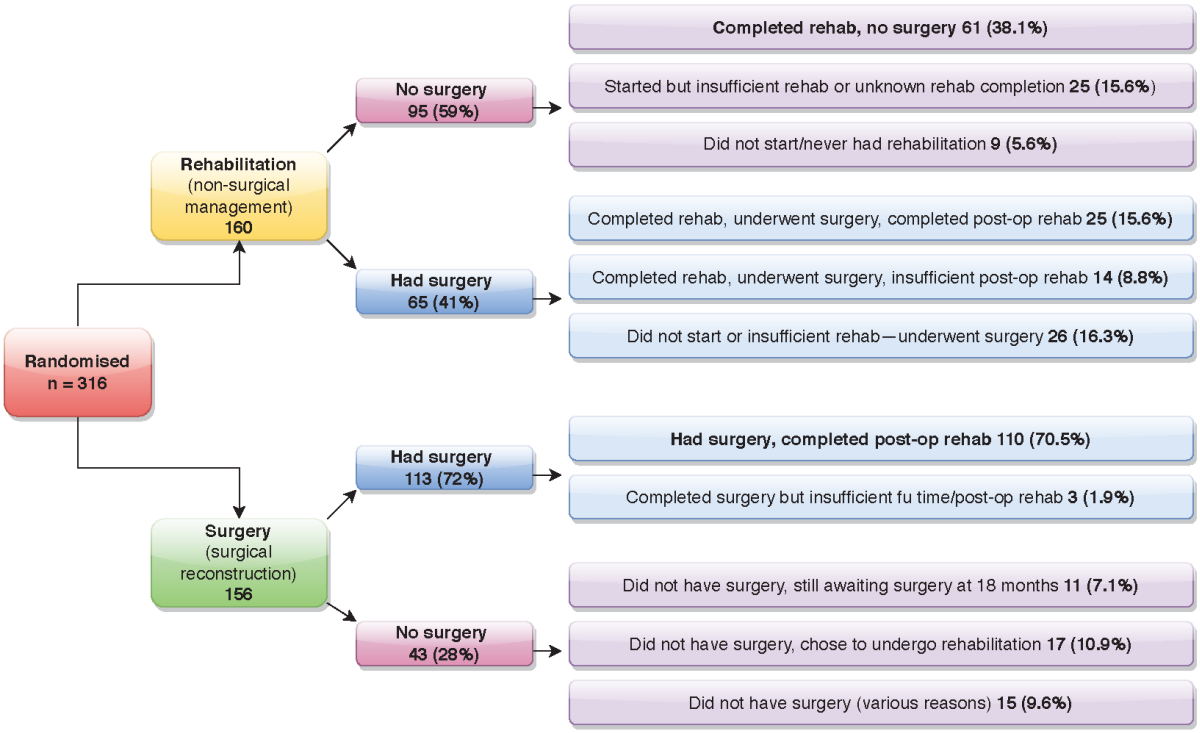



**Figure 5.3:** Adherence to treatment. Adapted from Beard et al., 2022[349]

Adherence was moderate in the non-surgical reconstruction arm, with 78% (125 of 160) patients randomized to the rehabilitation receiving any amount of physiotherapy (tables 5.6).

| ACL-SNNAP: Treatment adherence by intervention arm | |
|---|---|
| **Non-Surgical Management (Rehabilitation)** | **n (%)** |
| Completed rehabilitation, no surgery | 61 (38.1) |
| Completed rehabilitation but underwent surgery, completed post-op rehabilitation | 25 (15.6) |
| Completed rehabilitation but underwent surgery, insufficient post-op rehabilitation | 14 (8.8) |
| Started rehabilitation but insufficient rehabilitation or unknown rehabilitation completion | 25 (15.6) |
| Did not start rehabilitation (never had any rehabilitation) | 9 (5.6) |
| Did not start or insufficient rehabilitation—underwent surgery | 26 (16.3) |
| **Surgical Management (Reconstruction)** | **n (%)** |
| Had surgery, completed post-op rehabilitation | 110 (70.5) |
| Did not have surgery | 15 (9.6) |
| Did not have surgery, still awaiting surgery at 18 months | 11 (7.1) |
| Did not have surgery, underwent rehabilitation | 17 (10.9) |
| Completed surgery, but insufficient follow up time/post-operative rehabilitation | 3 (1.9) |

**Table 5.6:** Treatment adherence and non-adherence. Adapted from Beard et al., 2022[349]

There was moderate non-adherence to surgical reconstruction in the surgical arm, with only 113 of 156 (72%) of participants allocated to that arm receiving ACL reconstruction (tables 5.4 and 5.6). Among the 28% of participants who ultimately declined surgery, some opted out of reconstruction because their knee function improved before surgical reconstruction, others declined because of work related conflicts, while 7% of participants in the were still awaiting surgery at the end of the trial follow-up period. In 9.6% of cases of non-adherence, participants did not have surgery for unknown reasons. Among the 113 participants completing surgical reconstruction, adherence to rehabilitation (post-surgical) was also higher than in the non-surgical arm, with 110 of 156 patients (70.51%) completing rehabilitation according to the mandatory aims of the surgical arm compared to 59.4% among participants randomised to the rehabilitation arm.

Adherence to intervention allocation was higher in the rehabilitation arm (78.13%) than in the surgical arm (72%) (tables 5.4 and 5.6), but participants' adherence to the rehabilitation program

(CONSIDER Receipt) was lower, with 95 of the 160 participants allocated to the rehabilitation arm attending physiotherapy over a minimum timespan of 60 days, for an adherence rate of 59.4%. Thirty-four (24%) participants randomized to the non-surgical reconstruction arm had insufficient rehabilitation (less than the minimum 2 sessions over 60 days). Of those, 9 had no record of having started rehabilitation, while 25 started rehabilitations but it was unknown how much of the intervention they completed. These figures highlight both the difficulties of maintaining adherence to rehabilitation in real-world settings, and collecting and validating data for non-adherence from routinely collected records in pragmatic trials.[28,525]

A substantial percentage (41%) of participants allocated to the rehabilitation arm ultimately underwent surgery. This was an expected outcome for participants in the trial and may reflect limited ability to stabilize an unstable, ACL-deficient knee with exercise therapy alone rather than insufficient rehabilitation due to poor participant adherence.[519] Previous studies of rehabilitation versus surgical management for persons with acute ACL injury found 50% of participants went on to require surgery.[519] However, it is important to note that 26 (40%) of those participants in the rehabilitation arm, who ultimately underwent reconstruction surgery, did so early on in their course of treatment, having no or very little physiotherapy.

### What can this tell us about adherence in complex intervention clinical trials?

ACL-SNNAP also afforded an opportunity to explore factors influencing adherence, and non-adherence, to interventions in a live pragmatic trial. Various factors contributed to the adherence rate in both intervention arms, including the impact of the COVID-19 pandemic on

the availability of surgery and provision of physiotherapy.[349] One hundred twenty-six descriptions for participant non-adherence or participants dropping out of physiotherapy were extracted from the physiotherapists' CRFs. These were coded and thematically analysed with NVivo 12 to create seven themes for poor participant adherence or discharge from physiotherapy (table 5.6).

A key finding of this analysis is that the most frequently reported reason for non-adherence to rehabilitation were participants not being satisfied with their progress with rehabilitation or 26% (33/129) of reported reasons. This was followed by reinjury or flare-up of knee pain or instability for 16%, (21/129) of reports, difficulties with scheduling, securing childcare, getting to physiotherapy or moving to another city in 15% (19/129), and the COVID-19 pandemic in 9%, (11/129) of reported reasons for non-adherence. Participants having difficulty performing the rehabilitation exercises accounted for only 7% (9/129) of reported reasons, while participant illness accounted for 6% of reported causes of poor or non-adherence. The cause of poor or non-adherence was unknown in 20% (26/129) of cases. In the surgical reconstruction group, the reason for not undergoing surgery was the COVID-19 pandemic (14%), work commitments (7%), and participants moving out of the area (7%), while the reason for non-adherence was unknown in 32% of participants who were allocated to the surgical group but did not have reconstructive surgery.

| Theme | n | Examples of descriptive passage from case notes |
|---|---|---|
| **Pt. Scheduling difficulties** | **9** | |
| **Childcare** difficulties or other family-related difficulties made it difficult to attend therapy. | (3) | "Patient cancelled appointment on day due to childcare commitment." |
| Unable to keep schedule because of work or other commitments | (5) | "Unable to attend physiotherapy due to work commitments. patient was discharged from orthopaedic consultant clinic and physiotherapy department."<br>"Starting a new job with travel."<br>"busy with exams." |
| Patient late for appointment | (1) | "Patient late therefore only time for advice" |
| **Pt moved away** | **10** | |
| Patient moved away or abroad-or requested physio elsewhere | 10 | "Sessions stopped because patient moved abroad."<br>"Patient is moving abroad. Poor compliance with exercises."<br>"Patient relocated to London " (Also to Spain, Lithuania, Canada, and USA) |
| **Pt. difficulty with dosage or compliance with program** | **9** | |
| Patient found the frequency of physio difficult to maintain- time consuming | (5) | "Finding it difficult to stick to consistent paced rehab."<br>"Patient reports she has had little time to complete exercises as taking exams"<br>"overloaded and exercises regressed. Advice given on importance of consistent exercise" |
| Poor compliance with exercises | (4) | "Finding it difficult to stick to consistent paced rehab. Aware that surgery may not be an option."<br>"Patient hasn't improved at all with any of the markers and her compliance with her HEP is poor, she is unable to recall her exercises or demonstrate them effectively." |
| **Poor patient attendance-no reasons given** | **26** | |
| Patient was discharged because of poor attendance | (8) | "discharged from physiotherapy and class due to poor attendance."<br>"Discharged not attendance or contact for 4 months to the physiotherapy department"<br>"patient attended once and dna x5" |
| Patient stopped attending physio-no reason given | (18) | "Stopped attending the class, not responsive to attempts to contact."<br>"Participant did not return to see physiotherapist after 2nd treatment session."<br>"This patient was seen for an initial assessment and was given lower limb strengthening programme, but subsequently failed to attend any further sessions."<br>"During this time, I the Research Nurse and the physio tried to contact the patient to see if he was ok and if he wanted further appointments either with Physio or with a consultant. Despite a few calls and an email, no-one answered the calls, and the patient did not respond to the email I sent..." |
| **COVID-19 Pandemic Effects** | **11** | |
| **Covid-19** pandemic interrupted care | 11 | "ACL class was disrupted by Covid-19. Patient encourage for home exercise, details given to contact the department if feel she need physiotherapy after covid, but she has not contacted."<br>"Please note Physiotherapy sessions stopped early due to Covid pandemic. Physiotherapy clinic closed due to staff redeployment to Covid wards. Patient would normally receive 12 sessions." |

| | | |
|---|---|---|
| | | “Due to coronavirus done over the phone.” |
| Life event or unforeseen event | 1 | “Friend died so spoke on the phone.” |
| Re-injury or flare-up of pain or instability | 21 | |
| Patient was re-injured | (11) | “"Re-injury to Left knee Feb 2020. Second MRI shows partial tear to ACL graft and a tear to lateral meniscus. Awaiting arthroscopy. Physio D/C as no clinical benefit.”<br>“had further fall and new MRI showed bucket handle tear-not to be included in further study physio.” |
| Patient was experiencing pain-stopped or decreased compliance with physio | (10) | “In flare up -limited participation “<br>“Patient states increase pain. Struggling to come to terms with chronic knee pain.”<br>Patient reports ongoing pain and requested a review in clinic." |
| Illness | 8 | |
| Patient was ill or unwell | 8 | “Has not been able to do rehab has been unwell.”<br>“patient unwell tonsillitis.”<br>“recent appendectomy.” “Injured left side.” |
| Pt. unsatisfied with progress | 33 | |
| Patient felt knee was still too unstable | (12) | “Patient requested referral back to surgeon for instability.”<br>“Swapped from Physio-surgery as still complaining of instability “<br>“Knee continued to give way despite rehab. Referred back to Orthopaedic consultant.”<br>“Decision made patient was an ACL “non-coper” and listed for ACL reconstruction surgery.” |
| Patient not satisfied with progress, or elected surgery | (21) | “The pt requested a follow-up appointment with the surgeons but the physio was unaware of this. However, the patient then indicated that he was dissatisfied with his progress and wanted to seek a second opinion in Lithuania.”<br>“Patient had an appointment with a consultant and requested surgery due to being dissatisfied with physio progress.”<br>“Patient hasn't improved at all with any of the markers and her compliance with her HEP is poor, she is unable to recall her exercises or demonstrate them effectively. Still complains of pain and instability.”<br>“Patient had an appointment with a consultant and requested surgery due to being dissatisfied with physio progress.”<br>“reports he would like to have the ACL reconstructed and discontinue Physiotherapy.”<br>“Patient was not satisfied with their physio progress and wanted surgery.” |
| Patient satisfied with physio, but still preferred surgery | 2 | “Patient did not want to return to sport without having reconstruction due to increased risk of re-injury.” |
| Physio discharged patient: no clinical benefit | 1 | Physio D/C as no clinical benefit |

Table 5.6: Kinds of participant non-adherence and reasons

## AIM II: How do fidelity and adherence issues influence interpretation of the trial results?

The ACL-SNNAP trial found that surgical reconstruction was a clinically superior and more cost-effective management strategy for persons with non-acute ACL injury and persistent knee instability to rehabilitation management (table 5.7).[349] An adjusted mean difference (KOOS4) of 7.9 (95% CI 2.5, 13.2; p=0·0053) favoured of surgical management over rehabilitation. Intention to treat (ITT) analyses did not substantially alter this, resulting in an adjusted mean difference of 8.3 (3·3, 13.3),p=0·0012. However, participant adherence and intervention fidelity varied between both intervention arms, and the complex nature of the interventions created several potential treatment pathways through which participants progressed. The degree of participant adherence had a quantifiable influence on treatment effect estimates.

Both the PPc and PPp analyses supported the ITT analysis results, significantly favouring surgical management[349]. However, the PPc, representing the highest degree of adherence, resulted in adjusted and unadjusted mean differences that were lower than the primary outcome (ITT) analyses, while both the adjusted and unadjusted PPp analyses, which included any amount of adherence to treatment, resulted in larger treatment effects than the ITT (table 5.7).[349] Nevertheless, these pre-specified analyses undertaken to account for potential non-adherence consistently indicated a benefit for surgical management over rehabilitation for the trial's primary outcome (KOOS4) at 18 months. In a trial with large, robust treatment effects, like ACL-SNNAP, the per protocol analyses may ultimately not have been necessary. However, it can be difficult to predict, *a priori*, to predict the magnitude of treatment effects that will be found in a trial, and whether per-protocol analyses will influence their interpretation.

| Study arm (KOOS4 scores) | **Surgical reconstruction Mean (SD)** | **Rehabilitation Mean (SD)** | **Surgical reconstruction vs. rehabilitation** Mean difference (95% CI) | p value |
|---|---|---|---|---|
| **Mean, intention to treat (SD)** | | | | |
| Adjusted* | 72.4 (24.4) n=128 | 64.88 (22.41) n=120 | 7.86 (2.5-13.2) | 0.005 |
| Unadjusted | 72.97 (19.98) N=128 | 64.63 (19.98) N=120 | 8.34 (3.3-13.3) | 0.001 |
| **Per protocol conservative** | | | | |
| Adjusted | 76.1 (19.87) n=94 | 68.8 (20.15) n=73 | 7.3 (0.8-13.8) | 0.030 |
| Unadjusted | 75.9 (16.1) N=94 | 69.05 (17.26) N=73 | 6.85 (1.5-12.2) | 0.012 |
| **Per protocol pragmatic** | | | | |
| Adjusted | 75.83 (19.71) n=95 | 64.6 (22.04) n=100 | 11.23 (5.7-16.8) | < 0.001 |
| Unadjusted | 75.66 (19.07) N=95 | 64.76 (19.07) N=100 | 10.9 (5.5-16.3) | < 0.001 |
| *Main result for primary outcome | | | | |

**Table 5.7:** Intention to treat and per-protocol analyses for ACL-SNNAP primary outcome (KOOS4). Adapted from Beard et al., 2022[349]

In ACL-SNNAP, issues of adherence to allocated treatment and fidelity of treatment content were potential limitations, though threats from these factors were low. Despite detailed recording and reporting of adherence and fidelity in the rehabilitation arm, these were a relative weakness in the trial. Adherence data allow some confidence in determining whether participants in the surgical arm underwent their allocated treatment or not, but this was less distinct in the rehabilitation arm.[526] This could be perceived as a weakness in the design of the trial. However, it was also an expected pattern of adherence for a pragmatic trial of rehabilitation interventions. The nature of the two management strategies in ACL-SNNAP may have influenced this. While the surgical procedure was very clearly defined, rehabilitation was less defined and took place over a significantly larger length of time and treatment sessions.

## AIM III. What insights about the results of the meta-epidemiological study were gained through the experience of trial conduct and reporting in ACL-SNNAP ?

Per protocol analyses in ACL-SNNAP produced results echoing findings in the meta-epidemiological study in chapter IV. In the meta-epidemiological study, systematic differences were found between effect estimates in complex intervention trials achieving intervention fidelity and participant adherence compared to those without it. Intervention fidelity had a quantifiable influence on treatment effect estimates, with lower degrees of fidelity and adherence in the experimental intervention associated with larger treatment effect estimates (meta-regression coefficient 0.235, 95%CI: 0.675,-0.402).[526] In ACL-SNNAP, all treatment effects favoured surgical management, with both an adjusted mean difference (7.86 95% CI 2.54,13.19) and unadjusted mean difference (8.34 95%CI: 3.3, 13.3), or a moderate Cohen's d of 0.42. The PPP and PPc analyses supported the intention to treat analysis results. However, the PPc, representing the highest degree of adherence and fidelity, resulted in a slightly smaller mean difference in favour of surgical management than the ITT results (adjusted: 7.3 95% CI 0.8, 13.8 and unadjusted: 6.85, 95%CI: 1.5,12.2), or a smaller Cohen's d of 0.35.

However, while the treatment effect estimate was reduced by the conservative per protocol analysis (PPc – an analysis that was "fidelity strict"), it did not change the overall findings favouring surgical management over rehabilitation in ACL-SNNAP. As also found in the meta-epidemiological study, trials with moderate to large treatment effects ( > 0.40, Cohen's effect size[457]) were more robust against potential bias (SMD of -0.23 to -0.24) introduced by poor or absent fidelity than those with small pooled treatment effects (0.2 - 0.49[457]). This reinforces our

conclusion that poor or absent intervention fidelity can bias treatment effect estimates, but the implications should be interpreted in the context of individual trials, and with care.

## DISCUSSION

### CONSIDER in a pragmatic RCT: insights gained in the ACL-SNNAP trial

The primary aim of this chapter was to gather insights about the CONSIDER framework and the results of the meta-epidemiological study through the greater level of granularity available from data gathered prospectively in an ongoing clinical trial. In essence, we set out to "field test" the CONSIDER framework in the ACL-SNNAP trial. Key aspects of the framework and the meta-epidemiological study were validated and several important insights about fidelity monitoring and fidelity reporting in complex intervention clinical trials were gained.

The ACL-SNNAP trial presented challenges for fidelity, both in its design as a pragmatic trial comparing management strategies rather than stand-alone interventions, and in the unique circumstances resulting from conducting the trial during the COVID-19 pandemic. The wide range of interventions applied in the rehabilitation arm reflect routine clinical practice in the NHS but made it impossible to evaluate intervention fidelity at the level of individual participant sessions or interventions. The CONSIDER framework constructs fidelity in a multifaceted way, rather than modelling fidelity through one lens alone (for example, only as strictly controlled delivery of an intervention as per the treatment manual).

The CONSIDER framework facilitated identification of fidelity to the interventions arms' program theory, rather than individual treatments, better suiting the aims and context of the trial.[505] It also facilitated consideration of the contribution of both the interventionists' and the participants' actions (adherence to treatment allocation and compliance with the prescribed amount of treatment) to intervention fidelity. The framework's fidelity components, design (the program theory for the trial and goals of the management strategies), delivery (actions of the interventionists) and receipt (actions of the participants), facilitated fidelity monitoring that was tailored to a complex trial in a real-world setting.

This approach supports those employed in pragmatic trials of other complex interventions of patient care management strategies, such as the BRITE trial.[527] The BRITE study is a pragmatic trial investigating the effectiveness of two transitional care models for patients with traumatic brain injury (TBI) being discharged from inpatient rehabilitation. Both ACL-SNNAP and BRITE investigated complex management strategies, allowing flexible implementation based on patient and site-specific factors. The experience of applying CONSIDER in the ACL-SNNAP illustrated that, as in the BRITE study, intervention fidelity can be challenging to define and monitor in highly pragmatic trials. Such trials may require a flexible fidelity model that can be adapted to the aims and contexts of individual trials, rather than a rigidly defined framework assessing a single dimension of intervention fidelity.[216,527]

There were also limitations to CONSIDER's approach in this type of trial. Fidelity assessment frameworks are typically designed to rate set criteria for intervention delivery the same way

across all participant engagements with an intervention. However, this may not be possible, or even desirable, with pragmatic trials of intervention management strategies as in ACL-SNNAP. The focus of the trial was on the effectiveness of the management strategy, rather than its component interventions, encompassing a wide range of rehabilitation interventions and a high degree of allowable intervention tailoring. This reflects real-world physiotherapy practice and physiotherapy as an applied science, rather than a distinct treatment, but also increased the complexity of assessing fidelity in this kind of pragmatic trial.[332,524] Nevertheless, we were able to assess the fidelity of participants' intervention program to the trial's underpinning program theory (part of CONSIDER's design and delivery categories) and participant adherence (CONSIDER-receipt), reflecting the aims and objectives of ACL-SNNAP and the framework's flexible fidelity construct.

Other aspects of the design of the CONSIDER's framework were also explored through its use in ACL-SNNAP. During the best-fit framework synthesis used to formulate CONSIDER, participant enactment, or whether study participants use treatment-related skills in their daily lives, was excluded from the final synthesized fidelity framework.[311] We proposed that participant enactment would represent behaviours occurring in participants' everyday life, away from the trial setting, and there would be no way to observe or assess it in a clinic-based trial. We also theorised that a trial participant may receive a treatment delivered with perfect fidelity, and yet be unwilling or unable to apply it in daily life.[311] This may be influenced by a variety of factors not related to the degree of fidelity with which the intervention was delivered, such as low

intervention acceptability, forgetting to do it, lacking a suitable setting or time to perform the activities, or losing interest in the intervention.[51]

This rationale was supported by the thematic analysis of poor participant adherence in ACL-SNNAP. Around 70% of the recorded reasons for participants' poor or non-attendance in physiotherapy treatment sessions or their choosing to discontinue physiotherapy related to factors lying outside of the interventions (rehabilitation) themselves , such as participants being too busy, having scheduling difficulties, difficulties arranging childcare, or illness. Only 4% of the reasons for poor adherence related to the participants' difficulties performing exercises between therapy sessions or maintaining the frequency of interventions. A much larger percentage of reports of participant poor adherence (26%) were attributed to participants being unsatisfied with their progress with rehabilitation, supporting our previous suggestions that participant enactment may be an indication of the efficacy of interventions, rather than the fidelity with which they are delivered.[40]

The analyses of participant adherence data in this chapter also illustrated how a variety of factors can influence adherence to rehabilitation or surgery in pragmatic trials. Several other factors influencing adherence to physiotherapy had been identified previously, ranging from pre-injury activity levels, pain, depression, self-efficacy, social support, the duration of physiotherapy and complexity of therapeutic interventions the participants must perform.[28,528,529] The COVID-19 pandemic also introduced unique challenges to adherence in ACL-SNAPP, whether by reducing the availability of surgical reconstruction or the feasibility of in-person rehabilitation.

Despite these challenges, adherence to rehabilitation and surgery in ACL-SNNAP ranged between 59-63% and 71-72%, respectively. Lower participant adherence to outpatient physiotherapy for lower extremity and musculoskeletal conditions has previously been described, ranging from 10-35%, [28,528,530] while non-adherence can reach upwards of 70% in primarily self-managed, home-based physiotherapy programmes.[529]

An important finding in ACL-SNNAP was that the conservative per protocol analyses (PPc), representing the highest degree of fidelity and participant adherence, resulted in only a small reduction in treatment effect estimates derived from the ITT analysis. This further supports findings for the superiority of surgical management, indicating they were not likely due to poor fidelity in the rehabilitation arm. As found in the meta-epidemiological study, fidelity and adherence can have a quantifiable influence on treatment effect estimates. Like the results of the meta-epidemiological study, the size of the difference may not change the interpretation of treatment efficacy or effectiveness when trials result in robust estimates of treatment effects. Nevertheless, the potential difference in treatment effects influenced by intervention fidelity is important to consider. In a trial with smaller differences in treatment effect estimates between interventions than ACL-SNNAP, a similar bias in treatment effect estimates may lead to different interpretations about intervention effectiveness. This was demonstrated in the loss of statistical significance in pooled treatment effects found in 3 of the meta-analyses in Chapter IV when their treatment effect estimates were adjusted for a hypothetical loss of fidelity (table 5.7).

These findings highlight the importance of considering intervention fidelity and participant adherence, and how to analyse and interpret them, when analysing the results of pragmatic trials. They also support the use of per-protocol analyses, which rely on fidelity and adherence monitoring, to complement intention to treat (ITT) analyses in complex intervention RCTs. Intention-to-treat analyses estimate the effect assignment to an intervention in a trial, but not the effect of the treatment itself on participant outcomes.[99] They are agnostic about poor participant adherence, protocol deviations, use of proscribed co-interventions, participant withdrawal, and anything that happens after randomization.[98,99]

Intention to treat analyses will signal sufficiently large treatment effects (signals), but differences in treatment effects resulting from poor fidelity or adherence may bias their results. For example, a potential limitation identified in the ACL-SNNAP publication was that the due to the proportion of participants allocated to surgical management who did not undergo surgical reconstruction, the true benefit of surgical reconstruction could have been somewhat greater than suggested by the ITT analysis.[349]

As demonstrated in ACL-SNNAP, randomised controlled trials of interventions such as physiotherapy can produce complex clinical pathways that present challenges for fidelity and adherence. Including all participants randomized to a group in primary analyses, even if they were non-adherent, dropped out early, or received modified or poorly delivered treatment, may lead to under or overestimated treatment effects.[98,531] Poor fidelity or adherence concentrated in one study arm may also bias estimates of treatment effects toward one treatment, or diminish

differences between treatments if they occur across study arms.[98,531] For example, participants in a study with a rehabilitation treatment arm who did not receive sufficient therapy or stopped attending physiotherapy early could weaken the treatment effect of the rehabilitation treatment arm in an ITT analysis, biasing the analysis of the treatment's effectiveness. In ACL-SNNAP for example, improved adherence in the rehabilitation arm could lead to a lower mean difference between the surgical and rehabilitation arm, while better adherence in the surgical arm could increase the difference in outcomes between them.

Per-protocol (PP) analyses can complement ITT analyses by taking deviations from the study protocol and participants' adherence to treatment strategies into account, providing an estimate of the true efficacy of an intervention among those who completed the treatment as planned.[98,99] Per-protocol analyses can also become a valuable strategy for sensitivity analyses when a trial investigates interventions with convoluted clinical pathways or greater scope for variations in intervention delivery, such as the ACL-SNNAP trial. However, care must be taken when designing and interpreting PP analyses, as subjects may withdraw from a study or be non-adherent with treatment due to their response to the treatment.[98] Participants who do or do not adhere to a study protocol may also differ from each other with respect to pre-randomization and post-randomization prognostic factors that may have influenced their adherence, and this should be adjusted for in PP analysis.

Nevertheless, the PP effect in a trial is less likely to be biased by incomplete adherence or departures from the study protocol and may provide treatment effect estimates of greater

interest for patients or persons considering whether to use a treatment.[98,531] A pragmatic trial may ask, essentially, whether an intervention c*an work* in real-world settings, as well as whether it *does* work given how providers and participants interact with it. If a trial investigates whether an intervention *can work*, intervention fidelity and adherence data are essential for interpreting the results. For example, someone trying to decide whether to use a contraception method may want to know its estimated effectiveness when it was used as indicated, rather than the estimated effectiveness in a sample in which 40% of study participants did not use is properly.[98] If the trial asks whether an intervention *does work* in the context or "landscape" it is meant to be used in, a more pragmatic per-protocol approach may be needed (the PPp in ACL-SNNAP). Here, for example, fidelity and adherence data may provide important information about how the contraception worked when the consumer missed or mistimed one of the doses, but still used the contraception.

Given the substantial resources required to conduct pragmatic trials, complementing ITT effect estimates with PP effect estimates can provide a more comprehensive understanding of intervention effects and increase the amount of information available for clinical decision making. However, per protocol analyses cannot be undertaken if participant adherence and intervention fidelity are not monitored, further underscoring the importance of intervention fidelity in the interpretation of treatment effect estimates in clinical trials.

Insights were also gained into other aspects of CONSIDER recommendations. In line with the CONSIDER framework's recommended actions, surgical and rehabilitation arm case report forms

(CRFs) were disseminated and completed by interventionists participating in the trial with a high rate of adherence. Case report forms can offer several advantages for fidelity assessment. They can resemble clinical documentation typically used in routine care settings, can be unobtrusive, and avoid time-consuming documentation that may interrupt the intervention.[37,525] They can also offer systematic collection of intervention delivery details and contemporaneous records of participant responses when written within a reasonable time after intervention delivery.[37,525]

Case report forms have disadvantages as well, however, and those were experienced in this study. Case report forms may limit trialists' ability to evaluate fidelity in pragmatic trials if they do not collect intervention details needed for fidelity assessment. The ACL-SNNAP CRFs provided information about overall intervention content, allowing for evaluation of fidelity to the program theory, but granularity about fidelity at the individual intervention level was not possible. However, determining the relative effectiveness of the rehabilitation arms' component interventions was not the goal of the ACL-SNNAP trial.

Additionally, analyses of data from the CRFs in table 5.3 er pooled data from the CRFs for all across all sessions for all participants. This allowed for an overview of fidelity across all participants in the trial in aggregate but did not permit analysis of patterns for individual participants, or groups of participants, over individual sessions, over time. Individual variations in both fidelity and outcomes could not be assessed, but may yield valuable insights into participant characteristics influencing fidelity and adherence and variations in fidelity and adherence over time in individual participants. However, future secondary analyses of data from

ACL-SNNAP may explore these and yield further insights into fidelity and adherence in pragmatic complex intervention trials.

### What do fidelity monitoring and fidelity reporting in ACL-SNNAP tell us about the meta-epidemiological study in chapter IV?

A caveat in the meta-epidemiological study (chapter IV) is that reporting quality in RCT publications may not reflect the quality of trial conduct.[490] This raises risk of conflating the relationship between reporting of study conduct and treatment effect with that of actual conduct and treatment effect. It is possible that intervention fidelity was monitored but not reported in some of the RCTs assessed in chapter IV. Indeed, we found that randomised controlled trials with moderate to large effect sizes reported monitoring fidelity less often than trials with small effect sizes (Chapter IV, table 4.3). Similarly, the ACL-SNNAP trial publication in the Lancet reports a large treatment effect but fidelity and adherence were reported only briefly.

### Fidelity reporting: Is absence of proof a proof of absence?

The ACL-SNNAP publication in the Lancet describes monitoring of participant adherence, use of surgical case report forms to monitor compliance with intervention guidelines, recording of the content and adherence to post-surgical intervention, and the results of the per-protocol analyses.[349] However, given the recent experience of working in the ACL-SNNAP trial to assess intervention fidelity and participant adherence, it is clear that the amount of fidelity and

adherence monitoring and analysis undertaken in the trial was certainly greater than what was able to be included, or was essential to, the report for the trial publication, especially given the robust treatment effects found. This included collecting, coded and extensively analysing 1657 rehabilitation and surgical CRFs to gather important data about participant adherence and fidelity to the interventions' program theory. Data for participant adherence and intervention fidelity were also analysed extensively, as described in this chapter, guided by a theoretical framework (CONSIDER), addressing a limitation to fidelity monitoring reported in pragmatic trial and complex interventions literature.[224,497]

While fidelity may not have been monitored in papers not reporting it, ACL-SNNAP shows that it cannot be assumed with absolute certainty that not reporting it means it was not monitored at all. Editorial constraints limit the amount of information that can be published in papers in commercial publications, raising the possibility that other trials analysed in chapter IV also have conducted more extensive fidelity monitoring than was reported in their trial publications. Trials with robust treatment effect estimates, like ACL-SNAAP, may have monitored fidelity and participant adherence and conducted per-protocol analyses with fidelity data but not reported these due to word count and space limits, or because these were not essential data, given the trial's results. Fidelity data may have been included in online supplementary materials, appendices, or in grey literature publications rather than the main paper, raising the risk that they could be missed by keyword-based literature searches or while being screened for fidelity monitoring. For example, the National Institute for Health and Care Research (NIHR) Health Technology Assessment (HTA) Report for ACL-SNNAP[526] contained much greater detail about

fidelity and adherence than the primary trial publication in the Lancet, though this report may not have been found in a literature search in commercial literature search engines.

### Chicken or egg? Fidelity reporting and treatment effects

These factors create a "chicken or egg" conundrum, in which it becomes difficult to determine if trials' treatment effects influence fidelity reporting, or if fidelity influences reported treatment effects. While ACL-SNNAP cannot definitively answer or solve this conundrum, it highlights how reporting of fidelity can depend on several factors, including a trial's results, and is a major factor in the relationship between intervention fidelity and treatment effect estimates observed in chapter IV. An inescapable limitation of meta-epidemiological studies, including the one in this thesis, is that they assess the study characteristics that were reported in trial publications, rather than through direct observations of the trials themselves. The result of direct observation of intervention fidelity and participant adherence in the ACL-SNNAP trial show how the results of meta-epidemiological studies should be interpreted with prudent caution when reporting of the characteristic of interest is poor.

### What do CONSIDER and ACL-SNNAP tell us about intervention fidelity in pragmatic clinical trials?

The results of this chapter, and this thesis, make important contributions to the knowledgebase for intervention fidelity in complex intervention clinical trials. However, there continues to be much we do not yet fully understand about intervention fidelity, requiring further consideration

and investigation. There is little debate about the benefits of intervention fidelity in explanatory trials, but less consensus supporting fidelity monitoring in pragmatic clinical trials.[225,502,503] Trials, or their features, can fall along a continuum between being highly pragmatic or highly explanatory.[391] Intervention fidelity may need to be constructed and monitored differently in a trial, depending on where it falls along that continuum.

This was borne out in ACL-SNNAP, which required a different fidelity construct than would have been desirable for an explanatory trial. An explanatory trial which would have been more likely to favour greater emphasis on the delivery of individual interventions and their faithfulness to an intervention or study manual and monitoring provider competence (drift) over the course of the trial, for example. There may be an "outer limit" on the pragmatic trial side of the spectrum that includes explanatory and pragmatic trials, beyond which fidelity monitoring is no longer informative, feasible or desirable. That cannot be determined from the research conducted in this chapter, or this thesis, but presents a compelling question for exploration in future research.

## Future directions

Investigators in future pragmatic trials may build on our experiences in ACL-SNNAP when designing fidelity assessments for pRCTs. For more explanatory trials, specifying interventions core elements and components, *a priori*, are essential for fidelity monitoring. In highly pragmatic pRCTs or pRCTs of management strategies however, other targets for fidelity are needed. Alternate measures and markers of intervention fidelity can be identified and built-in to the trial

at the study protocol stage, helping researchers develop new essential targets for fidelity monitoring.[40,332,524] This will also facilitate decision-making over how fidelity is to be monitored and what, or how many, adaptations can be made to interventions without compromising their effectiveness.

Consistent monitoring and reporting of intervention fidelity is needed to determine whether adaptations to treatments or treatment programs during a trial impact their effectiveness or reproducibility.[532] These adaptations may be made for a variety of purposes and have differing implications. Some may enhance outcomes by adapting the intervention to the specific participants or context while preserving core elements needed for the intervention to be effective.[533] Other modifications may remove key intervention elements, impacting intervention's effects. It may be unknowable, *a priori*, how some modifications, such as adapting the dose or context of interventions, may influence participant outcomes. These may need to be investigated *post-hoc*, underscoring the importance of fidelity monitoring during the trial.

Trialists using CRFs to monitor fidelity may also devise and pilot forms that allow for more detailed intervention documentation without impeding the intervention in routine clinical practice or consider use of wearable devices or other methods for gathering data about participant adherence to trial activities. Future trials may also build opportunities for fidelity assessment directly into the intervention, reducing the need for more complex CRFs. Both objectives may also be supported by use of participant intervention logs (on paper or online), wearable devices, or other methods that create opportunities for study participants'

engagement with intervention activities to be recorded while yielding fidelity data that is less prone to recall or social desirability bias than directly asking intervention participants about their adherence with interventions.[525]

## CONCLUSION

The CONSIDER framework was initially designed for use in trials of distinct interventions, which have more defined interventions and allow for more detailed monitoring of intervention fidelity. Nevertheless, it provided a flexible structure for evaluating fidelity and adherence in a pragmatic trial comparing two patient management strategies with a large scope for variations in content and dosage. This application of the framework illustrated both its strengths and limitations. It also added important insights to the findings for the influence of fidelity on treatment effect estimates from the meta-epidemiological study in chapter IV. As demonstrated by ACL-SNNAP, published fidelity may not reflect actual in-trial fidelity. Reporting bias cannot be ruled out in meta-epidemiological studies.

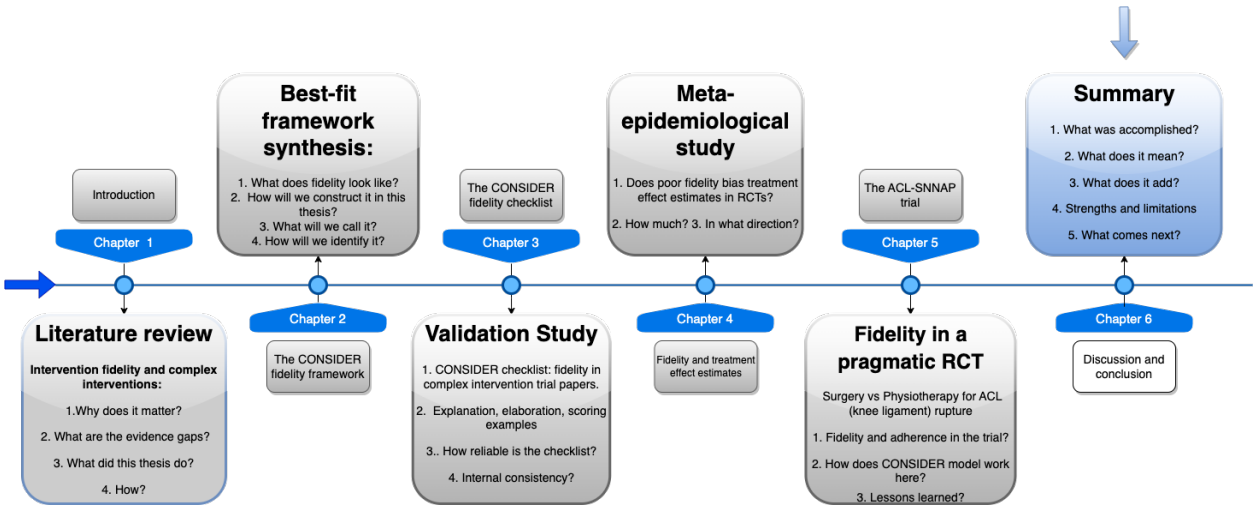
Fidelity monitoring in ACL-SNNAP also underscored the importance of incorporating fidelity and adherence data into trials' statistical analyses plans. Given the substantial resources required to plan and conduct pragmatic trials, monitoring fidelity and adherence and applying that data in per-protocol analyses can provide a more comprehensive understanding of intervention effects and increase the amount of information available for clinical decision making.

The results of the meta-epidemiological study, this chapter, and application of CONSIDER in ACL-SNNAP adds to the growing body of literature on fidelity assessment in complex intervention clinical trials. Pragmatic trials provide important evidence for clinical decision making, patients, and the public.[534] Ignoring intervention fidelity and participant adherence may increase the risk of accepting ineffective interventions and discarding potentially effective interventions that have been poorly implemented.

Chapter VI: Discussion and conclusions

## Chapter Summary

In this chapter, the aims and objectives achieved, insights gained, and strengths and limitations of this thesis are discussed. Future research directions and opportunities arising from this thesis are also presented.



This thesis investigated the effects of intervention fidelity on the results, interpretation, and appropriateness to change practice of clinical studies of complex interventions. It addressed several important knowledge gaps for intervention fidelity in complex intervention clinical trials. The results of this thesis also raise new questions about intervention fidelity that have important implications for future research on clinical trials in physical complex interventions.

## Aims achieved and their implications: Aims I and II

The lack of a uniform definition of fidelity and its key components has been identified as a barrier to fidelity planning and intervention delivery in clinical trials.[40,375,456] It may also act as a barrier to the translation of evidence-based interventions to clinical practice.[36,40,375] This thesis addressed this gap in physical complex interventions (rehabilitation) research by constructing an empirically based intervention fidelity framework, the **Co**mplex I**n**terventions De**si**gn, **D**elivery, **R**eceipt (CONSIDER) framework in chapter II, and field-testing it in the ACL-SNNAP trial in chapter V. It is, to the best of our knowledge, the first empirically derived fidelity framework created specifically for rehabilitation clinical trials.

The CONSIDER framework provides a platform for those who develop, implement, and study physical complex interventions to develop robust methods to support and assess intervention fidelity. It can also complement existing reporting guidelines such as SPIRIT[26], TIDieR[150], CERT[151], and IDEAL[165] by providing more targeted, detailed guidance for monitoring and reporting fidelity in physical complex interventions. Further development of the framework through exposure to a broader range of perspectives and consensus building can refine the framework and its

recommendations, facilitating practicable, evidence-based guidance for monitoring and maintaining intervention fidelity in complex intervention RCTs.

The large number of terms and descriptions used inconsistently to describe intervention fidelity in complex interventions literature also undermines identification of fidelity in trial publications.[1,12,40] This thesis addressed this gap by also developing a reliable, internally consistent, and feasible fidelity assessment checklist that distinguishes both trial actions supporting intervention fidelity as well as fidelity reported in trial publications (chapters II and III). Although the checklist was developed to facilitate transparent and reproducible fidelity assessment for this thesis (chapter IV), it can also have important applications more broadly. Fidelity assessment with the checklist can complement existing tools such as GRADE (Grading of Recommendations, Assessment, Development and Evaluations)[535] assessing certainty around the findings of systematic reviews, meta-analysis, and evidence-based practice guidelines for rehabilitation interventions. It can also be useful for grant reviewers, funding bodies, and journal editors evaluating treatment outcome studies and applications.

Clinicians and policymakers rely on the synthesis of the best available evidence to inform decision-making. Generally, systematic review and meta-analyses of data from high quality RCTs is the strongest source of evidence for intervention efficacy and effectiveness.[70] However, while much guidance has been developed for assessing the methodological quality of RCTs included in systematic reviews and meta-analyses, little attention has been focused on assessing the quality of the interventions themselves or the fidelity with which they were delivered. [2,101,472,483] Quality

assessment and risk of bias tools regularly used in complex interventions rarely include items for intervention fidelity or provide detailed guidance for assessing it.[71,487] For example, the most recent Cochrane risk of bias guidelines (ROB-2) includes potential bias arising from deviations from trials' intended interventions.[536] However, little detail is given to clarify what kind of deviations are intended or how they can be identified.[36,146,472,474] The CONSIDER checklist could be used, or developed further, to complement risk of bias and study quality assessments by providing more targeted assessment of intervention fidelity in trial publications.

## AIMS III and IV

This thesis provided the first empirical evidence of bias arising from poor intervention fidelity in the treatment effect estimates of clinical trials in rehabilitation in chapter IV. It had previously been proposed that poor intervention fidelity in clinical trials would diminish treatment effects due to contamination bias and extraneous variables.[53] The effect of poor intervention fidelity on clinical trial treatment effects was theorized however, rather than quantified, as presented in the thesis.

The central hypothesis of the thesis was that poor intervention fidelity produces a quantifiable bias in treatment effect estimates in complex in clinical trials. The findings of the meta-epidemiological study in chapter IV supported this hypothesis. There were systematic differences between effect estimates in complex intervention trials maintaining intervention fidelity compared to those without it. The estimated bias was robust to adjustment for several other

factors influencing trial outcomes and was of a magnitude that could influence interpretations of intervention effectiveness, particularly in trials with small to medium intervention effect estimates. This can have wide reaching implications in rehabilitation research, given that previous systematic reviews have found that rehabilitation RCTs often result in small treatment effect estimates.[537]

Contrary to the hypothesis, however, the meta-epidemiological study found evidence that poor or absent intervention fidelity resulted in larger (overestimated), rather than smaller (underestimated), treatment effect estimates in complex intervention RCTs. The direction of bias arising from poor intervention fidelity was an unexpected result, given predictions that poor intervention fidelity would dilute intervention effects and reduce trials' statistical power to detect them.[53,63,89,90] The mechanisms underlying an inverse relationship between intervention fidelity and intervention effect estimates are uncertain and likely complex. The overestimation of treatment effects may reflect the cumulative effects of introducing additional, unplanned interventions and active ingredients during intervention delivery[57,75,472–474] It may also reflect additional doses of the intended active ingredients if fidelity is not monitored.

It is also possible that intervention fidelity interacts with other design characteristics of complex intervention RCTs in unpredictable ways. For example, intervention fidelity may be associated with biased effect estimates to a greater degree in studies where outcome assessors were unblinded, or to a lesser degree in RCTs comparing an intervention to a wait-list control, rather than an active comparison. These and other possibilities could be explored in research building

of the findings of this thesis.

The direction of bias identified in chapter IV may also be influenced by factors within interventions or trials that cannot be determined through meta-epidemiological study. Intervention fidelity moderates intervention effects but does not produce them. The bias we estimated may also depend on a combination of the underlying soundness or "strength" of an intervention and the degree of modification it was subjected to.

A poorly conceived or ineffective intervention can be delivered with a high degree of intervention fidelity, but still be ineffective. An effective, well-conceived intervention may have been delivered with poor fidelity and still have resulted in robust treatment effects. Some types of interventions may be more susceptible to poor fidelity than others, or more responsive to changes in dosage or delivery. It is also possible that there is an element of selection bias in the meta-epidemiological study. It may be that trials that addressed fidelity well, and reported it, may be altogether more robust and less likely to show exaggerated treatment effects than less robust trials that also reported fidelity poorly, and were more likely to result in inflated treatment effects. This can also be explored in future work that builds on our results.

Nevertheless, the findings of this thesis show, empirically, that intervention fidelity is essential for accurate interpretation of intervention effects.[56] Like other characteristics associated with larger (overestimated) treatment effect sizes in RCTs, such as poor methodological quality[81], randomization[83], allocation concealment[82,85], or blinding,[78,84] fidelity does matter. Treatment

effect estimates biased by poor intervention fidelity could lead to the adoption of ineffective interventions, or premature rejection of interventions that could be effective. Bias arising from poor intervention fidelity can have important implications for evidence-based practice.

## Aim IV

This thesis also helps to address knowledge gaps about the feasibility, applicability, and impact of monitoring intervention fidelity in pragmatic RCTs. While there is little debate about the benefits of intervention fidelity in explanatory trials, less consensus supports fidelity monitoring in pragmatic clinical trials.[225,502,503] Effectiveness research can be challenging and comlplex.[538] How to assess fidelity feasibly, and to which intervention components, can be difficult to determine in pragmatic trials. Assessment of fidelity in ACL-SNNAP in chapter V showed that intervention fidelity may need to be modelled more flexibly or have different targets in a pragmatic RCT than in an explanatory trial. This thesis also showed that this flexible approach to fidelity is practicable.

This thesis also demonstrated how the nature or amount of fidelity monitoring appropriate for pragmatic trials may depend on the trial's goals and context (chapter V). Fidelity monitoring may exist along a continuum, in which moving the needle from more controlled trials to more highly pragmatic trials can shift the target of fidelity from higher magnification (greater focus) on individual intervention components to greater focus on the overall intervention package (lower magnification) that includes several factors surrounding the intervention itself. Future research

may also explore whether there is an “outer limit” on either end of the continuum between highly controlled explanatory trials and highly pragmatic trials beyond which fidelity monitoring is no longer feasible or informative.

## Limitations

The limitations of individual components of this thesis have been discussed in their respective chapters. Wider issues are briefly described here. First, it is important to acknowledge that the CONSIDER checklist and framework developed for this thesis are important first steps that need further development with broader input from a wider range of stakeholders. In future stages of their development, a Delphi process will be needed to build consensus about the definition of fidelity, which fidelity components and qualities are most important, and which qualities should be given the most weighting when developing and evaluating intervention fidelity in complex intervention trials.

Determining how often and how intervention fidelity is applied in complex intervention trials, and how fidelity may influence their results, inescapably depends on fidelity being reported in those trials. Reporting of Intervention details, including fidelity, is often poor in complex intervention trial publications. However, poor fidelity reporting may not reflect poor fidelity monitoring. Some papers may not have reported fidelity because of editorial constraints. It may also be the case that some rehabilitation papers did not report fidelity because it did not appear to influence the results or interpretation of trials with robust treatment effects. One of the most often cited reasons for assessing and reporting fidelity in psychology and behavioural health

trials is the need to account for negative or ambiguous findings.[221] In these cases, publication bias may also influence the amount of available published research reporting fidelity. Studies investigating the concordance between reported design characteristics and actual study conduct in rehabilitation research would be helpful in assessing the impact of this limitation for this thesis, as well as the larger body of meta-epidemiological research in complex interventions.

### Other future directions and opportunities arising from this thesis

This thesis addressed important evidence gaps for intervention fidelity in complex intervention clinical trials. Key aspects of intervention fidelity continue to need exploration and are outlined below.

### Control or comparison interventions

Primarily, this thesis examined the intervention fidelity of trials' experimental interventions. This was the focus of the thesis from the beginning. However, the in-depth analyses of intervention delivery reported in RCTs included in this thesis revealed that the control interventions in trials with an active intervention as a comparison were infrequently the subject of the same level of scrutiny or documentation as the experimental condition. Reporting of the experimental intervention details in complex intervention trials is generally poor.[539] This has been the subject of a number of systematic reviews was also found in this thesis. Control interventions are even more poorly described and reported than experimental interventions in complex intervention trials, however.[199]Systematic reviews of rehabilitation RCTs have found up to 75% of RCTs incompletely described the control intervention.[199,540] Other reviews have found that control

interventions also score significantly and consistently lower on the TIDieR checklist than experimental interventions.[199,540]

The intervention fidelity of control conditions has also been much less frequently assessed or reported.[139] Monitoring and assessing intervention fidelity in the experimental condition alone may not be enough to ensure trials have valid and reproducible results.[36] The extent to which a trial's control conditions included their core components also needs to be understood in order to establish causality between an intervention and outcomes.[139,541] This may be particularly important in pragmatic trials using "conventional therapy" or standard of care comparison groups.[139]

In many rehabilitation trials, the core components of the interventions being investigated may not be completely novel and may overlap with current practice or usual treatment.[19] The experimental interventions' core components or active ingredients may be present to some degree in the control condition, resulting in a loss of intervention differentiation and less contrast between conditions. This may lead to unpredictable, though possibly weaker, treatment effect estimates.[12,542] Despite this, the fidelity of the control conditions is often unmonitored or unreported, even in clinical trial papers reporting on intervention fidelity.

The true effect of poor fidelity on control conditions, relative to the experimental interventions, remains uncertain, however. To date, and to the best of our knowledge, meta-epidemiological studies have not examined the impact of poor fidelity during control intervention

implementation. Future research can build on the findings of this thesis, investigating how differences in degrees of intervention fidelity between the experimental and control conditions influence treatment effects in meta-epidemiological research.

Other descriptive and inferential analyses using fidelity data could be used to investigate the influence of fidelity differences between conditions for intervention outcomes. Between-group contrasts, or comparisons of fidelity levels in the experimental conditions, relative to levels of fidelity in the control group, could also be conducted by using or adapting already existing tools such as the Achieved Relative Strength Index.[542] Achieved relative strength is the difference between the experimental intervention, as implemented, and the control intervention, as implemented.[542] Low achieved relative treatment strength resulting from poor fidelity in either condition may limit the conclusions that can be drawn about intervention effectiveness.[542] This has not been shown empirically, however, presenting another opportunity for future research to build on the findings in this thesis.

### Intervention complexity and fidelity

The role of intervention complexity in the relationship between fidelity and intervention effects is another area in need of further research.[2] Interventions with more components and active ingredients may present more challenges for fidelity or be affected differently by varying degrees of intervention fidelity than more simple or straightforward interventions. Intervention complexity may act on a continuum along which levels of complexity may have a different effect on intervention fidelity.

Interventions with more components may present a greater scope for modification or poor delivery.[2,19] However, interventions with multiple components may also have overlapping active ingredients and may be more resilient to modified implementation than interventions with more consolidated mechanisms and ingredients. For example, an intervention targeting improved walking speed in persons recovering from stroke may include a combination of verbal cues and manual facilitation at the participant's pelvis and trunk. Decreased timing or precision in one set of cues may be overcome by quality of the effects of the other.

Intervention complexity may also influence the adherence of participants in a clinical trial.[537] For example, the surgical intervention in the ACL-SNNAP trial may be seen a more discrete, finite intervention than the rehabilitation intervention, which required a number of intervention sessions and a high degree of active participation in therapeutic activities by persons randomised to that arm. Future research may explore how varying degrees of intervention complexity influence participant adherence or interact with specific components of intervention fidelity (such as provider skill) to influence intervention outcomes, particularly in interventions taking place over many encounters or treatment sessions.

## CONCLUSION

In summary, the body of work contained in this thesis has met its goals of adding new knowledge and addressing evidence gaps for intervention fidelity in clinical trials of complex interventions. It estimated the prevalence of fidelity monitoring and reporting in complex intervention RCTs. It

also contributed an empirically based intervention fidelity framework and a reliable fidelity assessment checklist tailored to the unique needs and conditions experienced by researchers who conduct clinical trials in rehabilitation and those who seek to assess them. It also estimated empirically, for the first time, the magnitude and direction of bias in treatment effect estimates arising from poor intervention fidelity in rehabilitation clinical trials. It also explored intervention fidelity in an ongoing, contemporary pragmatic clinical trial, providing new insights into intervention fidelity and participant adherence in complex intervention pragmatic trials.

This thesis helps to address important gaps in our understanding of how intervention fidelity is monitored, supported, and influences the interpretation of intervention effects in rehabilitation research. It showed that intervention fidelity is important, not just on a theoretical, but also on an empirical level. The findings of this thesis provide new and important information for the conduct of clinical trials in rehabilitation and other complex interventions in physical domains of healthcare and the interpretation of their results.

## BIBLIOGRAPHY


1. McGee D, Lorencatto F, Matvienko-Sikar K, Toomey E. Surveying knowledge, practice and attitudes towards intervention fidelity within trials of complex healthcare interventions 11 Medical and Health Sciences 1117 Public Health and Health Services. Trials. 2018;19:1–14.

2. Carroll C, Patterson M, Wood S, Booth A, Rick J, Balain S. A conceptual framework for implementation fidelity. Implement Sci. 2007;2:1–9.

3. McCulloch P, Altman DG, Campbell WB, Flum DR, Glasziou P, Marshall JC, et al. No surgical innovation without evaluation: the IDEAL recommendations. Lancet. 2009;374:1105–12.

4. Hasson H. Systematic evaluation of implementation fidelity of complex interventions in health and social care. Implement Sci. 2010;5:67.

5. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ. 2008;a1655.

6. McCulloch P, Altman DG, Campbell WB, Flum DR, Glasziou P, Marshall JC, et al. No surgical innovation without evaluation: the IDEAL recommendations. Lancet. 2009;374:1105–12.

7. Beard D, Hamilton D, Davies L, Cook J, Hirst A, McCulloch P, et al. Evidence-Based Evaluation of Practice and Innovation in Physical Therapy Using the IDEAL-Physio Framework. Phys Ther. 2018;98:108–21.

8. Minary L, Trompette J, Kivits J, Cambon L, Tarquinio C, Alla F. Which design to evaluate complex interventions? Toward a methodological framework through a systematic review. BMC Med Res Methodol. 2019;19:1–9.

9. Beard DJ, Rees JL, Cook JA, Rombach I, Cooper C, Merritt N, et al. Arthroscopic subacromial decompression for subacromial shoulder pain (CSAW): a multicentre, pragmatic, parallel group, placebo-controlled, three-group, randomised surgical trial. Lancet. 2018;391:329–38.

10. Slaughter SE, Hill JN, Snelgrove-Clarke E. What is the extent and quality of documentation and reporting of fidelity to implementation strategies: A scoping review. Implement Sci. 2015;10:1–12.

11. Prowse PT. A Meta-Evaluation: The Role of Treatment Fidelity within Psychosocial Interventions during the Last Decade. J Psychiatry. 2015;18.

12. Toomey E, Hardeman W, Hankonen N, Byrne M, McSharry J, Matvienko-Sikar K, et al. Focusing on fidelity: narrative review and recommendations for improving intervention fidelity within trials of health behaviour change interventions. Heal Psychol Behav Med. 2020;8:132–51.

13. McCulloch P, Cook JA, Altman DG, Heneghan C, Diener MK. IDEAL framework for surgical innovation 1: the idea and development stages. BMJ. 2013;346:f3012–f3012.

14. Hawe P, Shiell A, Riley T. Theorising interventions as events in systems. Am J Community Psychol. 2009;43:267–76.

15. Hamilton AB, Moin T, Bean-Mayberry B, Finley EP, Huynh AK, Stirman SW, et al. A Pragmatic Approach to Guide Implementation Evaluation Research: Strategy Mapping for Complex Interventions. Front Public Heal. 2018;6.

16. Blencowe NS, Brown JM, Cook JA, Metcalfe C, Morton DG, Nicholl J, et al. Interventions in

randomised controlled trials in surgery: issues to consider during trial design. Trials. 2015;16:392.

17. Mak S, Hunt M, Boruff J, Zaccagnini M, Thomas A. Exploring professional identity in rehabilitation professions: a scoping review. Adv Heal Sci Educ. 2022;27:793–815.

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

19. Thibaut A, Beaudart C, Martens G, Bornheim S, Kaux JF. Common Bias and Challenges in Physical and Rehabilitation Medicine Research: How to Tackle Them. Front Rehabil Sci. 2022;3:1–12.

20. Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, et al. A new framework for developing and evaluating complex interventions: Update of Medical Research Council guidance. BMJ. 2021;374:1–11.

21. Cook CE, George SZ, Keefe F. Different interventions, same outcomes? Here are four good reasons. Br J Sports Med. 2018;52:951–2.

22. Haynes RB, Dantes R. Patient compliance and the conduct and interpretation of therapeutic trials. Control Clin Trials. 1987;8:12–9.

23. Persch AC, Page SJ. Protocol Development, Treatment Fidelity, Adherence to Treatment, and Quality Control. Am J Occup Ther. 2013;67:146–53.

24. Breckenridge A, Aronson JK, Blaschke TF, Hartman D, Peck CC, Vrijens B. Poor medication adherence in clinical trials: consequences and solutions. Nat Rev Drug Discov.

2017;16:149–50.

25. Dalli LL, Kilkenny MF, Arnet I, Sanfilippo FM, Cummings DM, Kapral MK, et al. Towards better reporting of the proportion of days covered method in cardiovascular medication adherence: A scoping review and new tool TEN-SPIDERS. Br J Clin Pharmacol. 2022;88:4427–42.

26. Chan AW, Tetzlaff JM, Gotzsche PC, Altman DG, Mann H, Berlin JA, et al. SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials. Bmj. 2013;346:e7586–e7586.

27. Katz JN, Losina E, Lohmander LS. OARSI Clinical Trials Recommendations: Design and conduct of clinical trials of surgical interventions for osteoarthritis. Osteoarthr Cartil. 2015;23:798–802.

28. Jack K, McLean SM, Moffett JK, Gardiner E. Barriers to treatment adherence in physiotherapy outpatient clinics: A systematic review. Man Ther. 2010;15:220–8.

29. Bassett S. Measuring Patient Adherence to Physiotherapy. J Nov Physiother. 2012;02.

30. Pullar T, Kumar S, Feely M. Compliance in clinical trials. Ann Rheum Dis. 1989;48:871–5.

31. Bollen JC, Dean SG, Siegert RJ, Howe TE, Goodwin VA. A systematic review of measures of self-reported adherence to unsupervised home-based rehabilitation exercise programmes, and their psychometric properties. BMJ Open. 2014;4:e005044.

32. Valgimigli M, Garcia-Garcia HM, Vrijens B, Vranckx P, McFadden EP, Costa F, et al. Standardized classification and framework for reporting, interpreting, and analysing medication non-adherence in cardiovascular clinical trials: a consensus report from the

Non-adherence Academic Research Consortium (NARC). Eur Heart J. 2019;40:2070–85.

33. Vrijens B, Urquhart J. Methods for Measuring, Enhancing, and Accounting for Medication Adherence in Clinical Trials. Clin Pharmacol Ther. 2014;95:617–26.

34. Corneli AL, Deese J, Wang M, Taylor D, Ahmed K, Agot K, et al. FEM-PrEP. JAIDS J Acquir Immune Defic Syndr. 2014;66:324–31.

35. Stinear CM, Lang CE, Zeiler S, Byblow WD. Advances and challenges in stroke rehabilitation. Lancet Neurol. 2020;19:348–60.

36. Pastva AM, Duncan PW, Reeves GR, Nelson MB, Whellan DJ, O'Connor CM, et al. Strategies for supporting intervention fidelity in the rehabilitation therapy in older acute heart failure patients (REHAB-HF) trial. Contemp Clin Trials. 2018;64:118–27.

37. Moncher FJ, Prinz RJ. Treatment fidelity in outcome studies. Clin Psychol Rev. 1991;11:247–66.

38. MEASURE Evaluation Implementation Research Technical Working Group. Fundamentals of Implementation. 2012;73.

39. Borrelli B, Sepinwall D, Ernst D, Bellg AJ, Czajkowski S, Breger R, et al. A new tool to assess treatment fidelity and evaluation of treatment fidelity across 10 years of health behavior research. J Consult Clin Psychol. 2005;73:852–60.

40. Gearing RE, El-Bassel N, Ghesquiere A, Baldwin S, Gillies J, Ngeow E. Major ingredients of fidelity: A review and scientific guide to improving quality of intervention research implementation. Clin Psychol Rev. 2011;31:79–88.

41. Ang K, Hepgul N, Gao W, Higginson IJ. Strategies used in improving and assessing the level

of reporting of implementation fidelity in randomised controlled trials of palliative care complex interventions: A systematic review. Palliat Med. 2018;32:500–16.

42. Parham LD, Cohn ES, Spitzer S, Koomar JA, Miller LJ, Burke JP, et al. Fidelity in sensory integration intervention research. Am J Occup Ther. 2007;61:216–27.

43. Zanca JM, Turkstra LS, Chen C, Packel A, Ferraro M, Hart T, et al. Advancing Rehabilitation Practice Through Improved Specification of Interventions. Arch Phys Med Rehabil. 2019;100:164–71.

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

45. Quested E, Ntoumanis N, Thøgersen-Ntoumani C, Hagger MS, Hancox JE. Evaluating quality of implementation in physical activity interventions based on theories of motivation: current challenges and future directions. Int Rev Sport Exerc Psychol. 2017;10:252–69.

46. Lambert JD, Greaves CJ, Farrand P, Cross R, Haase AM, Taylor AH. Assessment of fidelity in individual level behaviour change interventions promoting physical activity among adults: a systematic review. BMC Public Health. 2017;17:1–13.

47. Stirman SW, Miller CJ, Toder K, Calloway A. Development of a framework and coding system for modifications and adaptations of evidence-based interventions. Implement Sci. 2013;8:65.

48. Mars T, Ellard D, Carnes D, Homer K, Underwood M, Taylor SJC. Fidelity in complex behaviour change interventions: A standardised approach to evaluate intervention

integrity. BMJ Open. 2013;3:e003555.

49. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: Medical Research Council guidance. BMJ. 2015;350:h1258–h1258.

50. Ellefson MR, Oppenheimer DM. Is replication possible without fidelity? Psychol Methods. 2022;

51. Nelson DL, Mathiowetz V. Randomized controlled trials to investigate occupational therapy research questions. Am J Occup Ther. 2004;58:24–34.

52. Karas S, Plankis L. Consideration of treatment fidelity to improve manual therapy research. J Man Manip Ther. 2016;24:233–7.

53. Summerfelt WT. Program strength and fidelity in evaluation. Vol. 7, Applied Developmental Science. 2003. p. 55–61.

54. Paradis C. Bias in Surgical Research. Ann Surg. 2008;248:180–8.

55. Schwartz S, Carpenter KM. The right answer for the wrong question: Consequences of type III error for public health research. Am J Public Health. 1999;89:1175–80.

56. Perepletchikova F, Treat TA, Kazdin AE. Treatment Integrity in Psychotherapy Research: Analysis of the Studies and Examination of the Associated Factors. J Consult Clin Psychol. 2007;75:829–41.

57. Berkman ND, Lina Santaguida MP, Viswanathan M, Morton SC. The Empirical Evidence of Bias in Trials Measuring Treatment Differences. Rockville, Maryland; 2014.

58. Hawe P, Shiell A, Riley T. Complex interventions: how “out of control” can a randomised

controlled trial be? Bmj. 2004;328:1561–3.

59. Kaur P, Stoltzfus J. Type I, II, and III statistical errors: A brief overview. Int J Acad Med. 2017;3:268.

60. An M, Dusing SC, Harbourne RT, Sheridan SM. What Really Works in Intervention? Using Fidelity Measures to Support Optimal Outcomes. Phys Ther. 2020;100:757–65.

61. Dobson D, Cook TJ. Avoiding type III error in program evaluation. Eval Program Plann. 1980;3:269–76.

62. Salamh PA, Cory CS, Hanney WJ, Condon HE, Liu X, Kolber MJ. The Efficacy and Treatment Fidelity of Kinesiology Taping in Conjunction With Conservative Treatment Interventions Among Individuals With Shoulder Pain: A Systematic Review with Meta-Analysis. Int J Sports Phys Ther. 2021;16:606–19.

63. Ibrahim S, Sidani S. Fidelity of Intervention Implementation: A Review of Instruments. Health (Irvine Calif). 2015;07:1687–95.

64. Kinmonth AL, Wareham NJ, Hardeman W, Sutton S, Prevost AT, Fanshawe T, et al. Efficacy of a theory-based behavioural intervention to increase physical activity in an at-risk group in primary care (ProActive UK): a randomised trial. Lancet. 2008;371:41–8.

65. Hardeman W, Michie S, Fanshawe T, Prevost AT, Mcloughlin K, Kinmonth AL. Fidelity of delivery of a physical activity intervention: Predictors and consequences. Psychol Health. 2008;23:11–24.

66. Wareham NJ, Jakes RW, Rennie KL, Mitchell J, Hennings S, Day NE. Validity and repeatability of the EPIC-Norfolk Physical Activity Questionnaire. Int J Epidemiol.

2002;31:168–74.

67. Toomey E, Hardeman W. Addressing Intervention Fidelity Within Physical Therapy Research and Clinical Practice. J Orthop Sport Phys Ther. 2017;47:895–8.

68. Hassiotis A, Poppe M, Strydom A, Vickerstaff V, Hall IS, Crabtree J, et al. Clinical outcomes of staff training in positive behaviour support to reduce challenging behaviour in adults with intellectual disability: cluster randomised controlled trial. Br J Psychiatry. 2018;212:161–8.

69. Boruch RF, Gomez H. Sensitivity, bias, and theory in impact evaluations. Prof Psychol Res Pract. 1977;8:411–34.

70. Higgins JPT, Thomas J, Chandler J, Chumpston M, Li T, Page MJ, et al. Cochrane Handbook for Systematic Reviews of Interventions version 6.2 (updated February 2021). Vol. 6.2, The Cochrane Collaboration. Chichester, UK: John Wiley & Sons, Ltd; 2021.

71. Viswanathan M, Patnode CD, Berkman ND, Bass EB, Chang S, Hartling L, et al. Assessing the Risk of Bias in Systematic Reviews of Health Care Interventions. Agency for Healthcare Research and Quality Methods Guide for Effectiveness and Comparative Effectiveness Reviews. 2012.

72. Cook C, Garcia AN. Post-randomization bias. J Man Manip Ther. 2020;28:69–71.

73. Kunz R, Vist G, Oxman AD. Randomisation to protect against selection bias in healthcare trials. Cochrane database Syst Rev. 2007;MR000012--MR000012.

74. Bennett MI, Hughes N, Johnson MI. Methodological quality in randomised controlled trials of transcutaneous electric nerve stimulation for pain: Low fidelity may explain negative

findings. Pain. 2011;152:1226–32.

75. Hettema J, Steele J, Miller WR. Motivational Interviewing. Annu Rev Clin Psychol. 2005;1:91–111.

76. Walsh M, Srinathan SK, McAuley DF, Mrkobrada M, Levine O, Ribic C, et al. The statistical significance of randomized controlled trial results is frequently fragile: a case for a Fragility Index. J Clin Epidemiol. 2014;67:622–8.

77. Garcia-Alamino JM, Bankhead C, Heneghan C, Pidduck N, Perera R. Impact of heterogeneity and effect size on the estimation of the optimal information size: Analysis of recently published meta-analyses. BMJ Open. 2017;7:1–8.

78. Savović J, Jones HE, Altman DG, Harris RJ, Jüni P, Pildal J, et al. Influence of Reported Study Design Characteristics on Intervention Effect Estimates From Randomized, Controlled Trials. Ann Intern Med. 2012;157:429.

79. Sterne JACC, Jüni P, Schulz KF, Altman DG, Bartlett C, Egger M. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. Stat Med. 2002;21:1513–24.

80. Moustgaard H, Jones HE, Savović J, Clayton GL, Sterne JAC, Higgins JPT, et al. Ten questions to consider when interpreting results of a meta-epidemiological study—the MetaBLIND study as a case. Res Synth Methods. 2020;11:260–74.

81. Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? Lancet. 1998;352:609–13.

82. Page MJ, Higgins JPT, Clayton G, Sterne JAC, Hróbjartsson A, Savović J. Empirical Evidence of Study Design Biases in Randomized Trials: Systematic Review of Meta-Epidemiological Studies. Scherer RW, editor. PLoS One. 2016;11:e0159267.

83. Pildal J, Hrobjartsson A, Jorgensen K, Hilden J, Altman D, Gotzsche P. Impact of allocation concealment on conclusions drawn from meta-analyses of randomized trials. Int J Epidemiol. 2007;36:847–57.

84. Saltaji H, Armijo-Olivo S, Cummings GG, Amin M, da Costa BR, Flores-Mir C. Influence of blinding on treatment effect size estimate in randomized controlled trials of oral health interventions. BMC Med Res Methodol. 2018;18:42.

85. Armijo-Olivo S, Saltaji H, da Costa BR, Fuentes J, Ha C, Cummings GG. What is the influence of randomisation sequence generation and allocation concealment on treatment effects of physical therapy trials? A meta-epidemiological study. BMJ Open. 2015;5:e008562.

86. Dechartres A, Ravaud P, Atal I, Riveros C, Boutron I. Association between trial registration and treatment effect estimates: a meta-epidemiological study. BMC Med. 2016;14:100.

87. Saltaji H, Armijo-Olivo S, Cummings GG, Amin M, da Costa BR, Flores-Mir C. Impact of Selection Bias on Treatment Effect Size Estimates in Randomized Trials of Oral Health Interventions: A Meta-epidemiological Study. J Dent Res. 2018;97:5–13.

88. Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. BMJ. 2003;326:1167–70.

89. Dumas JE, Lynch AM, Laughlin JE, Phillips Smith E, Prinz RJ. Promoting intervention fidelity.

Conceptual issues, methods, and preliminary results from the EARLY ALLIANCE prevention trial. Am J Prev Med. 2001;20:38–47.

90. Robinson NB, Fremes S, Hameed I, Rahouma M, Weidenmann V, Demetres M, et al. Characteristics of Randomized Clinical Trials in Surgery From 2008 to 2020. JAMA Netw Open. 2021;4:1–13.

91. Robiner WN. Enhancing adherence in clinical research. Contemp Clin Trials. 2005;26:59–77.

92. Shiovitz TM, Bain EE, McCann DJ, Skolnick P, Laughren T, Hanina A, et al. Mitigating the Effects of Nonadherence in Clinical Trials. J Clin Pharmacol. 2016;56:1151–64.

93. Freedman LS. The effect of partial noncompliance of the power of a cancer prevention clinical trial. Control Clin Trials. 1990;11:269–70.

94. Steinberg WJ, Price M. Power and the Factors Affecting It. In: Statistics Alive! 3rd ed. SAGE Publications; 2020. p. 405–15.

95. Lachin JM. Introduction to sample size determination and power analysis for clinical trials. Control Clin Trials. 1981;2:93–113.

96. Pledger GW. Compliance in clinical trials: impact on design, analysis and interpretation. Epilepsy Res Suppl. 1988;1:125–33.

97. McIntyre A, Richardson M, Janzen S, Hussein N, Teasell R. The Evolution of Stroke Rehabilitation Randomized Controlled Trials. Int J Stroke. 2014;9:789–92.

98. Hernán MA, Robins JM. Per-Protocol Analyses of Pragmatic Trials. N Engl J Med. 2017;377:1391–8.

99. Fisher LD, Dixon D, Herson J, Frankowski RK, Hearron MS, Peace KE. Intention to treat in clinical trials. In: Peace KE, editor. Statistical issues in drug research and development. New York: Marcel Dekker; 1990. p. 331–59.

100. May GS, DeMets DL, Friedman LM, Furberg C, Passamani E. The randomized clinical trial: bias in analysis. Circulation. 1981;64:669–73.

101. Roen K, Arai L, Roberts H, Popay J. Extending systematic reviews to include evidence on implementation: Methodological work on a review of community-based initiatives to prevent injuries. Soc Sci Med. 2006;63:1060–71.

102. Altman DG. The Revised CONSORT Statement for Reporting Randomized Trials: Explanation and Elaboration. Ann Intern Med. 2001;134:663.

103. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJJ, et al. CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. Int J Surg. 2012;63:e1–37.

104. Balk EM. Correlation of Quality Measures With Estimates of Treatment Effect in Meta-analyses of Randomized Controlled Trials. JAMA. 2002;287:2973.

105. McGee D, Lorencatto F, Matvienko-Sikar K, Toomey E. Surveying knowledge, practice and attitudes towards intervention fidelity within trials of complex healthcare interventions. Trials. 2018;19:504.

106. Kyngäs H, Duffy ME, Kroll T. Conceptual analysis of compliance. J Clin Nurs. 2000;9:5–12.

107. Snowden A, Martin C, Mathers B, Donnell A. Concordance: a concept analysis. J Adv Nurs. Snowden A. 2014;70:46–59.

108. Gardner CL. Adherence: A Concept Analysis. Int J Nurs Knowl. 2015;26:96–101.

109. Avery L, Sniehotta FF, Denton SJ, Steen N, McColl E, Taylor R, et al. Movement as Medicine for Type 2 Diabetes: Protocol for an open pilot study and external pilot clustered randomised controlled trial to assess acceptability, feasibility and fidelity of a multifaceted behavioural intervention targeting physical activity in. Trials. 2014;15:1–18.

110. Scott SE, Breckon JD, Copeland RJ. An integrated motivational interviewing and cognitive-behavioural intervention promoting physical activity maintenance for adults with chronic health conditions: A feasibility study. Chronic Illn. 2018;174239531876937.

111. Merriam-Webster. (n.d.) [Internet]. 10th ed. Merriam-Webster.com dictionary. Springfield, Mass.: Merriam-Webster; [cited 2023 Jan 18]. Available from: https://www.merriam-webster.com/dictionary/

112. Online Etymology Dictionary [Internet]. Available from: https://www.etymonline.com

113. Alikari V, Zyga S. Conceptual analysis of patient compliance in treatment. Heal Sci J. 2014;8:179–86.

114. Felzmann H. Adherence, compliance, and concordance: an ethical perspective. Nurse Prescr. 2012;10:406–11.

115. Organization WH. Adherence to long-term therapies: evidence for action. World Health Organization; 2003.

116. Song MK, Happ MB, Sandelowski M. Development of a tool to assess fidelity to a psycho-educational intervention. J Adv Nurs. 2010;66:673–82.

117. Feely M, Seay KD, Lanier P, Auslander W, Kohl PL. Measuring Fidelity in Research Studies:

A Field Guide to Developing a Comprehensive Fidelity Measurement System. Child Adolesc Soc Work J. 2018;35:139–52.

118. Quinn L, Trubey R, Gobat N, Dawes H, Edwards RT, Jones C, et al. Development and Delivery of a Physical Activity Intervention for People With Huntington Disease: Facilitating Translation to Clinical Practice. J Neurol Phys Ther. 2016;40:71–80.

119. Lewis C, Darnell D, Kerns S, Monroe-DeVita M, Landes SJ, Lyon AR, et al. Proceedings of the 3rd Biennial Conference of the Society for Implementation Research Collaboration (SIRC) 2015: advancing efficient methodologies through community partnerships and team science. Implement Sci. 2016;11:85.

120. Eysenck HJ. The effects of psychotherapy: an evaluation. J Consult Psychol. 1952;16:319–24.

121. Rogers CR, Dymond RF. Psychotherapy and personality change. Rogers CR, Dymond RF, editors. Psychotherapy and personality change. Chicago, IL, US: University of Chicago Press; 1954.

122. Quay HC. The Three Faces of Evaluation: What can be Expected to Work. Crim Justice Behav. 1977;4:341–54.

123. Bergin AE, Lambert MJ. The evaluation of therapeutic outcomes. In: Handbook of psychotherapy and behavior change : an empirical analysis. 2nd ed. New York: Wiley; 1978. p. 139–89.

124. Cook TD, Campbell DT. Quasi-experimentation : design and analysis issues for field settings. Chicago: Rand McNally College; 1979.

125. Yeaton WH, Sechrest L. Critical dimensions in the choice and maintenance of successful treatments: Strength, integrity, and effectiveness. J Consult Clin Psychol. 1981;49:156–67.

126. Peterson L, Homer AL, Wonderlich SA. The integrity of independent variables in behavior analysis. J Appl Behav Anal. 1982;15:1308296.

127. Billingsley F, White OR, Munson R. Procedural reliability: A rationale and an example. Behav Assess. 1980;2:229–41.

128. Blakely CH, Mayer JP, Gottschalk RG, Schmitt N, Davidson WS, Roitman DB, et al. The fidelity-adaptation debate: Implications for the implementation of public sector social programs. Am J Community Psychol. 1987;15:253–68.

129. Brekke JS, Wolkon GH. Monitoring Program Implementation in Community Mental Health Settings. Eval Health Prof. 1988;11:425–40.

130. Lysynchuk LM, Pressley M, D'Ailly H, Smith M, Cake H. A Methodological Analysis of Experimental Studies of Comprehension Strategy Instruction. Read Res Q. 1989;24:458.

131. Meyer A, Miller S, Herman M. Balancing the priorities of evaluation with the priorities of the setting: A focus on positive youth development programs in school settings. J Prim Prev. 1993;14:95–113.

132. Bond GR, Drake RE. Assessing the Fidelity of Evidence-Based Practices: History and Current Status of a Standardized Measurement Methodology. Adm Policy Ment Heal Ment Heal Serv Res. 2020;47:874–84.

133. Bellg AJ, Resnick B, Minicucci DS, Ogedegbe G, Ernst D, Borrelli B, et al. Enhancing treatment fidelity in health behavior change studies: Best practices and recommendations

from the NIH Behavior Change Consortium. Heal Psychol. 2004;23:443–51.

134. Donabedian A. The Criteria and Standards of Quality. Health Administration Press; 1982. (Explorations in quality assessment and monitoring).

135. Lichstein KL, Riedel BW, Grieve R. Fair tests of clinical trials: A treatment implementation model. Adv Behav Res Ther. 1994;16:1–29.

136. Dane A V, Schneider BH. Program integrity in primary and earliy secondary prevention: Are implementation effects out of control? Clin Psychol Rev. 1998;18:23–45.

137. McGrew JH, Bond GR, Dietzen L, Salyers M. Measuring the Fidelity of Implementation of a Mental Health Program Model. J Consult Clin Psychol. 1994;62:670–8.

138. Borrelli B. The assessment, monitoring, and enhancement of treatment fidelity in publich health clinical trials. J Public Health Dent. 2012;71:S52–S63.

139. Nelson MC, Cordray DS, Hulleman CS, Darrow CL, Sommer EC. A Procedure for Assessing Intervention Fidelity in Experiments Testing Educational and Behavioral Interventions. J Behav Health Serv Res. 2012;39:374–96.

140. Bauer MS, Damschroder L, Hagedorn H, Smith J, Kilbourne AM. An introduction to implementation science for the non-specialist. BMC Psychol. 2015;3:1–12.

141. U.S. National Institutes of Health D of CC and PS (DCCPS) at the NCI (NCI). About Implementation Science [Internet]. [cited 2023 May 1]. Available from: https://cancercontrol.cancer.gov/is/about

142. Walton H, Spector A, Tombor I, Michie S. Measures of fidelity of delivery of, and engagement with, complex, face-to-face health behaviour change interventions: A

systematic review of measure quality. Br J Health Psychol. 2017;22:872–903.

143. Hand BN, Darragh AR, Persch AC. Thoroughness and psychometrics of fidelity measures in occupational and physical therapy: A systematic review. Am J Occup Ther. 2018;72:7205205050p1–10.

144. Brogan E, Ciccone N, Godecke E. A review of treatment fidelity literature in aphasia research. 2018.

145. Anis L, Benzies KM, Ewashen C, Hart MJ, Letourneau N. Fidelity Assessment Checklist Development for Community Nursing Research in Early Childhood. Front Public Heal. 2021;9:582950.

146. Pérez MC, Minoyan N, Ridde V, Sylvestre MP, Johri M. Comparison of registered and published intervention fidelity assessment in cluster randomised trials of public health interventions in low- and middle-income countries: systematic review protocol. Syst Rev. 2016;5:177.

147. Des Jarlais DC, Lyles C, Crepaz N. Improving the Reporting Quality of Nonrandomized Evaluations of Behavioral and Public Health Interventions: The TREND Statement. Am J Public Health. 2004;94:361–6.

148. Schulz KF, Altman DG, Moher D, Group C. CONSORT 2010 Statement : Updated Guidelines for Reporting Parallel Group Randomized Trials TO CONSORT. Ann Intern Med Acad Clin. 2010;152:1–8.

149. Boutron I, Altman DG, Moher D, Schulz KF, Ravaud P. CONSORT Statement for Randomized Trials of Nonpharmacologic Treatments: A 2017 Update and a CONSORT

Extension for Nonpharmacologic Trial Abstracts. Ann Intern Med. 2017;167:40.

150. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. Bmj. 2014;348:g1687–g1687.

151. Slade SC, Dionne CE, Underwood M, Buchbinder R, Beck B, Bennell K, et al. Consensus on Exercise Reporting Template (CERT): Modified Delphi Study. Phys Ther. 2016;96:1514–24.

152. Ginsburg LR, Hoben M, Easterbrook A, Anderson RA, Estabrooks CA, Norton PG. Fidelity is not easy! Challenges and guidelines for assessing fidelity in complex interventions. Trials. 2021;22:372.

153. Toomey E, Matvienko-Sikar K, Heary C, Delaney L, Queally M, Hayes CB, et al. Intervention Fidelity Within Trials of Infant Feeding Behavioral Interventions to Prevent Childhood Obesity: A Systematic Review. Ann Behav Med. 2019;53:75–97.

154. Schinckus L, Van den Broucke S, Housiaux M. Assessment of implementation fidelity in diabetes self-management education programs: A systematic review. Patient Educ Couns. 2014;96:13–21.

155. JaKa MM, Haapala JL, Trapl ES, Kunin-Batson AS, Olson-Bullis BA, Heerman WJ, et al. Reporting of treatment fidelity in behavioural paediatric obesity intervention trials: a systematic review. Obes Rev. 2016;17:1287–300.

156. Capin P, Walker MA, Vaughn S, Wanzek J. Examining How Treatment Fidelity Is Supported, Measured, and Reported in K–3 Reading Intervention Research. Educ Psychol Rev. 2018;30:885–919.

157. Kechter A, Amaro H, Black DS. Reporting of Treatment Fidelity in Mindfulness-Based Intervention Trials: A Review and New Tool Using NIH Behavior Change Consortium Guidelines. Mindfulness (N Y). 2019;10:215–33.

158. Keller RG, Ong AA, Nguyen SA, O'Connell BP, Lambert PR. Postinflammatory medial canal fibrosis: An institutional review and meta-analysis of short- and long-term outcomes. Laryngoscope. 2017;127:488–95.

159. Campbell M. Framework for design and evaluation of complex interventions to improve health. BMJ. 2000;321:694–6.

160. Barkun JS, Aronson JK, Feldman LS, Maddern GJ, Strasberg SM, Collaboration B. Surgical Innovation and Evaluation 1 Evaluation and stages of surgical innovations. 2009;1089–96.

161. Ergina PL, Cook JA, Blazeby JM, Boutron I, Clavien P alain, Reeves BC, et al. Surgical Innovation and Evaluation 2 Challenges in evaluating surgical innovation. Lancet. 2009;374:1097–104.

162. Sedrakyan A, Campbell B, Merino JG, Kuntz R, Hirst A, McCulloch P. IDEAL-D: A rational framework for evaluating and regulating the use of medical devices. BMJ. 2016;2372:i2372.

163. Verkooijen HM, Kerkmeijer LGW, Fuller CD, Huddart R, Faivre-Finn C, Verheij M, et al. R-IDEAL: A Framework for Systematic Clinical Evaluation of Technical Innovations in Radiation Oncology. Front Oncol. 2017;7.

164. Hirst A, Philippou Y, Blazeby J, Campbell B, Campbell M, Feinberg J, et al. No Surgical Innovation Without Evaluation: Evolution and Further Development of the IDEAL

Framework and Recommendations. Ann Surg. 2019;269:211–20.

165. Bilbro NA, Hirst A, Paez A, Vasey B, Pufulete M, Sedrakyan A, et al. The IDEAL Reporting Guidelines. Ann Surg. 2021;273:82–5.

166. Agha RA, Franchi T, Sohrabi C, Mathew G, Kerwan A, Thoma A, et al. The SCARE 2020 Guideline: Updating Consensus Surgical CAse REport (SCARE) Guidelines. Int J Surg. 2020;84:226–30.

167. Agha RA, Fowler AJ, Saeta A, Barai I, Rajmohan S, Orgill DP, et al. The SCARE Statement: Consensus-based surgical case report guidelines. Int J Surg. 2016;34:180–6.

168. COMET initiative [Internet]. [cited 2016 Sep 15]. Available from: http://www.comet-initiative.org

169. Blencowe NS, Mills N, Cook JA, Donovan JL, Rogers CA, Whiting P, et al. Standardizing and monitoring the delivery of surgical interventions in randomized clinical trials. Br J Surg. 2016;103:1377–84.

170. Fitzgerald GK, Hinman RS, Zeni J, Risberg MA, Snyder-Mackler L, Bennell KL. OARSI Clinical Trials Recommendations: Design and conduct of clinical trials of rehabilitation interventions for osteoarthritis. Osteoarthr Cartil. 2015;23:803–14.

171. Toomey E, Currie-Murphy L, Matthews J, Hurley DA. The effectiveness of physiotherapist-delivered group education and exercise interventions to promote self-management for people with osteoarthritis and chronic low back pain: A rapid review Part I. Man Ther. 2015;20:265–86.

172. Toomey E, Currie-Murphy L, Matthews J, Hurley DA. Implementation fidelity of

physiotherapist-delivered group education and exercise interventions to promote self-management in people with osteoarthritis and chronic low back pain: a rapid review part II. Man Ther. 2015;20:287–94.

173. Morgan C, Darrah J, Gordon AM, Harbourne R, Spittle A, Johnson R, et al. Effectiveness of motor interventions in infants with cerebral palsy: a systematic review. Dev Med Child Neurol. 2016;58:900–9.

174. Vaughan-Graham J, Cott C, Wright FV. The Bobath (NDT) concept in adult neurological rehabilitation: what is the state of the knowledge? A scoping review. Part II: intervention studies perspectives. Disabil Rehabil. 2015;37:1909–28.

175. Hafsteinsdóttir T, Kappelle J, Grypdonck M, Algra A. Effects of bobath-based therapy on depression, shoulder pain and health-related quality of life in patients after stroke. J Rehabil Med. 2007;39:627–32.

176. Page SJ, Schmid A, Harris JE. Optimizing terminology for stroke motor rehabilitation: recommendations from the American Congress of Rehabilitation Medicine Stroke Movement Interventions Subcommittee. Arch Phys Med Rehabil. 2012;93:1395–9.

177. Walker MF, Hoffmann TC, Brady MC, Dean CM, Eng JJ, Farrin AJ, et al. Improving the development, monitoring and reporting of stroke rehabilitation research: Consensus-based core recommendations from the Stroke Recovery and Rehabilitation Roundtable. Int J Stroke. 2017;12:472–9.

178. Heidlage JK, Cunningham JE, Kaiser AP, Trivette CM, Barton EE, Frey JR, et al. The effects of parent-implemented language interventions on child linguistic outcomes: A meta-

analysis. Early Child Res Q. 2020;50:6–23.

179. Nye C, Vanryckeghem M, Schwartz JB, Herder C, Turner HM, Howard C. Behavioral stuttering interventions for children and adolescents: A systematic review and meta-analysis. J Speech, Lang Hear Res. 2013;56:921–32.

180. Cleave PL, Becker SD, Curran MK, Van Horne AJO, Fey ME. The efficacy of recasts in language intervention: a systematic review and meta-analysis. Am J speech-language Pathol. 2015;24:237–55.

181. Brogan E, Ciccone N, Godecke E. Treatment fidelity in aphasia randomised controlled trials. Aphasiology. 2019;33:759–79.

182. Karas S, Plankis L. Treatment fidelity: Implications for manual therapy research. 2016.

183. Lim S, Hill B, Teede HJ, Moran LJ, O'Reilly S. An evaluation of the impact of lifestyle interventions on body weight in postpartum women: A systematic review and meta-analysis. Obes Rev. 2020;21:1–12.

184. Howlett N, Trivedi D, Troop NA, Chater AM. Are physical activity interventions for healthy inactive adults effective in promoting behavior change and maintenance, and which behavior change techniques are effective? A systematic review and meta-analysis. Transl Behav Med. 2019;9:147–57.

185. Scott JM, Zabor EC, Schwitzer E, Koelwyn GJ, Adams SC, Nilsen TS, et al. Efficacy of Exercise Therapy on Cardiorespiratory Fitness in Patients With Cancer: A Systematic Review and Meta-Analysis. J Clin Oncol. 2018;36:2297–305.

186. Ward TJC, Plumptre CD, Dolmage TE, Jones A V., Trethewey R, Divall P, et al. Change in VO

2peak in Response to Aerobic Exercise Training and the Relationship With Exercise Prescription in People With COPD: A Systematic Review and Meta-Analysis. Chest. 2020;158:131–44.

187. Hislop AC, Collins NJ, Tucker K, Deasy M, Semciw AI. Does adding hip exercises to quadriceps exercises result in superior outcomes in pain, function and quality of life for people with knee osteoarthritis? A systematic review and meta-analysis. Br J Sports Med. 2020;54:263–71.

188. Kunstler BE, Cook JL, Freene N, Finch CF, Kemp JL, O'Halloran PD, et al. Physiotherapist-Led Physical Activity Interventions Are Efficacious at Increasing Physical Activity Levels: A Systematic Review and Meta-analysis. Clin J Sport Med. 2018;28:304–15.

189. Depiazzi JE, Forbes RA, Gibson N, Smith NL, Wilson AC, Boyd RN, et al. The effect of aquatic high-intensity interval training on aerobic performance, strength and body composition in a non-athletic population: systematic review and meta-analysis. Clin Rehabil. 2019;33:157–70.

190. Salamh PA, Liu X, Hanney WJ, Sprague PA, Kolber MJ. The efficacy and fidelity of clinical interventions used to reduce posterior shoulder tightness: a systematic review with meta-analysis. J Shoulder Elb Surg. 2019;28:1204–13.

191. Salamh P, Cook C, Reiman MP, Sheets C. Treatment effectiveness and fidelity of manual therapy to the knee: A systematic review and meta-analysis. Musculoskeletal Care. 2017;15:238–48.

192. Hacke C, Nunan D, Weisser B. Do Exercise Trials for Hypertension Adequately Report

Interventions? A Reporting Quality Study. Int J Sports Med. 2018;39:902–8.

193. O'Shea O, McCormick R, Bradley JM, O'Neill B. Fidelity review: a scoping review of the methods used to evaluate treatment fidelity in behavioural change interventions. Phys Ther Rev. 2016;21:207–14.

194. Hildebrand MW, Host HH, Binder EF, Carpenter B, Freedland KE, Morrow-Howell N, et al. Measuring treatment fidelity in a rehabilitation intervention study. Am J Phys Med Rehabil. 2012;91:715–24.

195. Breckenridge J, Jones D. Evaluation in everyday occupational therapy practice: Should we be thinking about treatment fidelity? Br J Occup Ther. 2015;78:331–3.

196. Leeden M van der, Bart Staal J, Beekman E, Hendriks E, Mesters I, Rooij M de, et al. Development of a framework to describe goals and content of exercise interventions in physical therapy: a mixed method approach including a systematic review. Phys Ther Rev. 2014;19:1–14.

197. Yamato TP, Maher CG, Saragiotto BT, Hoffmann TC, Moseley AM. How completely are physiotherapy interventions described in reports of randomised trials? Physiother (United Kingdom). 2016;102:121–6.

198. Burgess LC, Wainwright TW, James KA, von Heideken J, Iversen MD. The quality of intervention reporting in trials of therapeutic exercise for hip osteoarthritis: a secondary analysis of a systematic review. Trials. 2021;22:1–11.

199. Lohse KR, Pathania A, Wegman R, Boyd LA, Lang CE. On the Reporting of Experimental and Control Therapies in Stroke Rehabilitation Trials : A Systematic Review. Arch Phys Med

Rehabil. 2018;99:1424–32.

200. Yu AM, Balasubramanaiam B, Offringa M, Kelly LE. Reporting of interventions and "standard of care" control arms in pediatric clinical trials: a quantitative analysis. Pediatr Res. 2018;84:393–8.

201. 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.

202. McCambridge AB, Nasser AM, Mehta P, Stubbs PW, Verhagen AP. Has reporting on physical therapy interventions improved in 2 decades? an analysis of 140 trials reporting on 225 interventions. J Orthop Sports Phys Ther. 2021;51:503–9.

203. Armijo-Olivo S, Cummings GG, Fuentes J, Saltaji H, Ha C, Chisholm A, et al. Identifying Items to Assess Methodological Quality in Physical Therapy Trials: A Factor Analysis. Phys Ther. 2014;94:1272–84.

204. Hart T, Dijkers MP, Whyte J, Turkstra LS, Zanca JM, Packel A, et al. A Theory-Driven System for the Specification of Rehabilitation Treatments. Arch Phys Med Rehabil. 2018;0.

205. Horne M, Thomas N, Vail A, Selles R, McCabe C, Tyson S. Staff's views on delivering patient-led therapy during inpatient stroke rehabilitation: a focus group study with lessons for trial fidelity. Trials. 2015;16:137.

206. Cohen DJ, Crabtree BF, Etz RS, Balasubramanian BA, Donahue KE, Leviton LC, et al. Fidelity Versus Flexibility. Am J Prev Med. 2008;35:S381–9.

207. Bird ML, Mortenson WB, Eng JJ. Evaluation and facilitation of intervention fidelity in

community exercise programs through an adaptation of the TIDier framework. BMC Health Serv Res. 2020;20:1–7.

208. Thorpe KE, Zwarenstein M, Oxman AD, Treweek S, Furberg CD, Altman DG, et al. A pragmatic–explanatory continuum indicator summary (PRECIS): a tool to help trial designers. J Clin Epidemiol. 2009;62:464–75.

209. Luce BR, Drummond M, Jönsson B, Neumann PJ, Schwartz J. S, Siebert U, et al. EBM, HTA, and CER: Clearing the Confusion. Milbank Q. 2010;88:256–76.

210. Keefe FJ, Jensen MP, Williams AC d. C, George SZ. The yin and yang of pragmatic clinical trials of behavioral interventions for chronic pain. Pain. 2021;Publish Ah.

211. Patsopoulos NA. A pragmatic view on pragmatic trials. Dialogues Clin Neurosci. 2011;13:217–24.

212. Dal-Ré R, Janiaud P, Ioannidis JPA. Real-world evidence: How pragmatic are randomized controlled trials labeled as pragmatic? BMC Med. 2018;16:49.

213. Toomey E, Matthews J, Guerin S, Hurley DA. Development of a Feasible Implementation Fidelity Protocol Within a Complex Physical Therapy–Led Self-Management Intervention. Phys Ther. 2016;96:1287–98.

214. Greenhalgh T, Papoutsi C. Studying complexity in health services research: desperately seeking an overdue paradigm shift. BMC Med. 2018;16:95.

215. Pinto RZ, Ferreira ML, Oliveira VC, Franco MR, Adams R, Maher CG, et al. Patient-centred communication is associated with positive therapeutic alliance: a systematic review. J Physiother. 2012;58:77–87.

216. Mignogna J, Martin LA, Harik J, Hundt NE, Kauth M, Naik AD, et al. “I had to somehow still be flexible”: exploring adaptations during implementation of brief cognitive behavioral therapy in primary care. Implement Sci. 2018;13:76.

217. Haynes A, Brennan S, Redman S, Williamson A, Gallego G, Butow P. Figuring out fidelity: a worked example of the methods used to identify, critique and revise the essential elements of a contextualised intervention in health policy agencies. Implement Sci. 2015;11:23.

218. Green LW. Making research relevant: if it is an evidence-based practice, where’s the practice-based evidence? Fam Pract. 2008;25:i20–4.

219. Schoenwald SK, Garland AF. A review of treatment adherence measurement methods. Psychol Assess. 2013;25:146–56.

220. Riley BL, MacDonald J, Mansi O, Kothari A, Kurtz D, vonTettenborn LI, et al. Is reporting on interventions a weak link in understanding how and why they work? A preliminary exploration using community heart health exemplars. Implement Sci. 2008;3:27.

221. Mowbray CT, Holter MC, Teague GB, Bybee D, Gregory B, Bybee D. Fidelity Criteria: Development, Measurement, and Validation. Am J Eval. 2003;24:315–40.

222. Zoldan Gonzalez G, Moseley AM, Maher CG, Nascimento DP, da Cunha Menezes Costa L, Oliveira Pena Costa L. Quality of musculoskeletal trials Methodological quality and statistical reporting of physical therapy randomized controlled trials relevant to musculoskeletal conditions. Arch Phys Med Rehabil. 2017;in press.

223. Jüni P, Altman DG, Egger M, Bmj S, Medical B, Jul N, et al. Systematic Reviews In Health

Care : Assessing The Quality Of Controlled Clinical Trials Linked references are available on JSTOR for this article : reviews in health Assessing the quality of controlled clinical trials other. 2015;323:42–6.

224. Vernooij-Dassen M, Moniz-Cook E. Raising the standard of applied dementia care research: Addressing the implementation error. Aging Ment Health. 2014;18:809–14.

225. Nieuwenhuis JB, Irving E, Oude Rengerink K, Lloyd E, Goetz I, Grobbee DE, et al. Pragmatic trial design elements showed a different impact on trial interpretation and feasibility than explanatory elements. J Clin Epidemiol. 2016;77:95–100.

226. Behn N, Harrison M, Brady MC, Breitenstein C, Carragher M, Fridriksson J, et al. Developing, monitoring, and reporting of fidelity in aphasia trials: core recommendations from the collaboration of aphasia trialists (CATs) trials for aphasia panel. Aphasiology. 2022;00:1–23.

227. Akobeng AK. Principles of evidence based medicine. Arch Dis Child. 2005;90:837–40.

228. Rixon L, Baron J, McGale N, Lorencatto F, Francis J, Davies A. Methods used to address fidelity of receipt in health intervention research: a citation analysis and systematic review. BMC Health Serv Res. 2016;16:663.

229. Nielsen B, Tse T, Haslam B, Carey LM. Development of an audit checklist to evaluate treatment fidelity of a complex rehabilitation intervention. Disabil Rehabil. 2022;0:1–8.

230. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 2009;6:e1000097.

231. Tong A, Flemming K, McInnes E, Oliver S, Craig J. Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. BMC Med Res Methodol. 2012;12:1–8.

232. Booth A, Clarke M, Dooley G, Ghersi D, Moher D, Petticrew M, et al. The nuts and bolts of PROSPERO: an international prospective register of systematic reviews. Syst Rev. 2012;1:2.

233. Carroll C, Booth A, Leaviss J, Rick J. " Best fit " framework synthesis : refining the method. 2013;

234. Booth A, Carroll C. Systematic searching for theory to inform systematic reviews : is it feasible ? Is it desirable ? 2015;220–35.

235. Booth A. Searching for qualitative research for inclusion in systematic reviews: a structured methodological review. Syst Rev. 2016;5:74.

236. Carroll C, Booth A, Cooper K. A worked example of best fit framework synthesis a systematic review of views concerning the taking. BMC Med Res Methodol. 2011;11.

237. Painter JE, Borba CPC, Hynes M, Mays D, Glanz K. The use of theory in health behavior research from 2000 to 2005: A systematic review. Annals of Behavioral Medicine. 2008.

238. United States National Library of Medicine. ClinicalTrials.gov [Internet]. Available from: https://www.clinicaltrials.gov

239. Paez A. Gray literature: An important resource in systematic reviews. J Evid Based Med. 2017;10:233–40.

240. Gardner T, Refshauge K, Smith L, McAuley J, Hübscher M, Goodall S. Physiotherapists' beliefs and attitudes influence clinical practice in chronic low back pain: a systematic

review of quantitative and qualitative studies. J Physiother. 2017;63:132–43.

241. Ames H, Glenton C, Lewin S. Purposive sampling in a qualitative evidence synthesis: A worked example from a synthesis on parental perceptions of vaccination communication. BMC Med Res Methodol. 2019;19:1–9.

242. Suri H. Purposeful Sampling in Qualitative Research Synthesis. Qual Res J. 2011;11:63–75.

243. Centre for Reviews and Dissemination U of Y. Systematic reviews: CRD's guidance for undertaking reviews in health care. Centre for Reviews and Dissemination, University of York, 2008. 2009. 294 p.

244. Booth A, Carroll C. How to build up the actionable knowledge base: the role of 'best fit' framework synthesis for studies of improvement in healthcare. BMJ Qual Saf. 2015;24:700–8.

245. Saldana J. The Coding Manual for Qualitative Researchers. 3rd ed. London: Sage Publications Inc.; 2016.

246. Wolcott HF. Transforming qualitative data : description, analysis, and interpretation. Thousand Oaks, CA: Sage Publications; 1994. 55 p.

247. Miles MB, Huberman AM. Qualitative Data Analysis: An Expanded Sourcebook. 2nd ed. Thousand Oaks, CA: Sage Publications; 1994. 69 p.

248. Nowell LS, Norris JM, White DE, Moules NJ. Thematic Analysis. Int J Qual Methods. 2017;16:160940691773384.

249. Miles MB, Huberman AM, Saldana J. Qualitative Data Analysis: A Methods Sourcebook. Third Edition. The SAGE Handbook of Applied Social Research Methods. 2014.

250. Barnett-Page E, Thomas J. Methods for the synthesis of qualitative research: a critical review. BMC Med Res Methodol. 2009;9:59.

251. Anderson CA, Bushman BJ, Bandura A, Braun V, Clarke V, Bussey K, et al. Using thematic analysis in psychology Using thematic analysis in psychology. Psychiatr Q. 2014;0887:37–41.

252. Sihvonen R, Paavola M, Malmivaara A, Itälä A, Joukainen A, Nurmi H, et al. Arthroscopic Partial Meniscectomy versus Sham Surgery for a Degenerative Meniscal Tear. N Engl J Med. 2013;369:2515–24.

253. Swank DJ, Swank-Bordewijk SCG, Hop WCJ, Van Erp WFM, Janssen IMC, Bonjer HJ, et al. Laparoscopic adhesiolysis in patients with chronic abdominal pain: A blinded randomised controlled multi-centre trial. Lancet. 2003;361:1247–51.

254. Thakur Y, Jones DL, Skanes A, Yee R, Drangova M. Right-Side RF Ablation Using Remote Catheter Navigation: Experimental Results In Vivo. J Cardiovasc Electrophysiol. 2012;23:81–7.

255. Sprowson AP, Jensen CD, Parsons N, Partington P, Emmerson K, Carluke I, et al. The effect of triclosan coated sutures on rate of Surgical Site Infection after hip and knee replacement: A protocol for a double-blind randomised controlled trial. 2014.

256. Sprowson AP, Jensen CD, Gupta S, Parsons N, Murty AN, Jones SM, et al. The effect of high dose antibiotic impregnated cement on rate of surgical site infection after hip hemiarthroplasty for fractured neck of femur: A protocol for a double-blind quasi randomised controlled trial. 2013.

257. Casida JM, Yaremchuk KL, Shpakoff L, Marrocco A, Babicz G, Yarandi H. The effects of guided imagery on sleep and inflammatory response in cardiac surgery: A pilot randomized controlled trial. J Cardiovasc Surg (Torino). 2013;54:269–79.

258. Weston M, Batterham AM, Tew GA, Kothmann E, Kerr K, Nawaz S, et al. Patients Awaiting Surgical Repair for Large Abdominal Aortic Aneurysms Can Exercise at Moderate to Hard Intensities with a Low Risk of Adverse Events. Front Physiol. 2017;7:1–11.

259. Stephens TJ, Peden CJ, Pearse RM, Shaw SE, Abbott TEF, Jones EL, et al. Improving care at scale: process evaluation of a multi-component quality improvement intervention to reduce mortality after emergency abdominal surgery (EPOCH trial). Implement Sci. 2018;13:142.

260. Wright N, Abantanga F, Amoah M, Appeadu-Mensah W, Bokhary Z, Bvulani B, et al. Developing and implementing an interventional bundle to reduce mortality from gastroschisis in low-resource settings. Wellcome open Res. 2019;4:46.

261. Cousins S, Blencowe NS, Tsang C, Lorenc A, Chalmers K, Carr AJ, et al. Reporting of key methodological issues in placebo-controlled trials of surgery needs improvement: a systematic review. J Clin Epidemiol. 2020;119:109–16.

262. Beard DJ, Campbell MK, Blazeby JM, Carr AJ, Weijer C, Cuthbertson BH, et al. Considerations and methods for placebo controls in surgical trials (ASPIRE guidelines). Lancet. 2020;395:828–38.

263. Cronbach LJ, Meehl P. The construct validity of psychological tests. Psychol Bull. 1955;52:281–301.

264. Dejong G, Horn SD, Gassaway JA, Slavin MD, Dijkers MP. Toward a taxonomy of rehabilitation interventions: Using an inductive approach to examine the 'black box' of rehabilitation. Arch Phys Med Rehabil. 2004;85:678–86.

265. Hurley DA, Hall AM, Currie-Murphy L, Pincus T, Kamper S, Maher C, et al. Theory-driven group-based complex intervention to support self-management of osteoarthritis and low back pain in primary care physiotherapy: Protocol for a cluster randomised controlled feasibility trial (SOLAS). BMJ Open. 2016;6.

266. Kaderavek JN, Justice LM. Fidelity: An Essential Component of Evidence-Based Practice in Speech-Language Pathology. Am J Speech-Language Pathol. 2010;19:369–79.

267. Behrman AL, Bowden MG, Rose DK. Clinical trials in neurorehabilitation. Handb Clin Neurol. 2013;110:61–6.

268. Carpenter JS, Burns DS, Wu J, Yu M, Ryker K, Tallman E, et al. Strategies Used and Data Obtained During Treatment Fidelity Monitoring. Nurs Res. 2013;62:59–65.

269. Dunn W, Little LM, Pope E, Wallisch A. Establishing Fidelity of Occupational Performance Coaching. OTJR (Thorofare N J). 2018;38:96–104.

270. Miller J, Barber D, Donnelly C, French S, Green M, Hill J, et al. PROPEL: implementation of an evidence based pelvic floor muscle training intervention for women with pelvic organ prolapse: a realist evaluation and outcomes study protocol. BMJ Open. 2017;7:843.

271. Pérez D, Van der Stuyft P, Zabala MC, Castro M, Lefèvre P. A modified theoretical framework to assess implementation fidelity of adaptive public health interventions. Implement Sci. 2016;11:1–11.

272. Bronars CA, Hanza MM, Meiers SJ, Patten CA, Clark MM, Nigon JA, et al. Treatment Fidelity Among Family Health Promoters Delivering a Physical Activity and Nutrition Intervention to Immigrant and Refugee Families. Heal Educ Behav. 2017;44:262–70.

273. Resnick B, Michael K, Shaughnessy M, Nahm ES, Sorkin JD, Macko R. Exercise intervention research in stroke: optimizing outcomes through treatment fidelity. Top Stroke Rehabil. 2011;18 Suppl 1:611–9.

274. Adams C, Lockton E, Gaile J, Earl G, Freed J. Implementation of a manualized communication intervention for school-aged children with pragmatic and social communication needs in a randomized controlled trial: the Social Communication Intervention Project. Int J Lang Commun Disord. 2012;47:245–56.

275. Bryant C, Lewis P, Bennell KL, Ahamed Y, Crough D, Jull GA, et al. Can physical therapists deliver a pain coping skills program? an examination of training processes and outcomes. Phys Ther. 2014;94:1443–54.

276. Forster A, Hartley S, Barnard L, Ozer S, Hardicre N, Crocker T, et al. An intervention to support stroke survivors and their carers in the longer term (LoTS2Care): study protocol for a cluster randomised controlled feasibility trial. Trials. 2018;19:317.

277. Campbell R, Rawlins E, Wells S, Kipping RR, Chittleborough CR, Peters TJ, et al. Intervention fidelity in a school-based diet and physical activity intervention in the UK: Active for Life Year 5. Int J Behav Nutr Phys Act. 2015;12:141.

278. Silveira SL, McCroskey J, Wingo BC, Motl RW. eHealth-Based Behavioral Intervention for Increasing Physical Activity in Persons With Multiple Sclerosis: Fidelity Protocol for a

Randomized Controlled Trial. JMIR Res Protoc. 2019;8:e12319.

279. Toomey E, Matthews J, Hurley DA. Using mixed methods to assess fidelity of delivery and its influencing factors in a complex self-management intervention for people with osteoarthritis and low back pain. BMJ Open. 2017;7:1–14.

280. Boden I, El-Ansary D, Zalucki N, Robertson IK, Browning L, Skinner EH, et al. Physiotherapy education and training prior to upper abdominal surgery is memorable and has high treatment fidelity: a nested mixed-methods randomised-controlled study. 2018.

281. Muir A, Ziviani J, Boyd R, Sakzewski L. Treatment fidelity in provider-supported parent-implemented intervention studies for infants with/at risk of developmental delay or disability: Systematic review. 2018.

282. Broekhuizen K, van Poppel MNM, Koppes LL, Kindt I, Brug J, van Mechelen W. Can Multiple Lifestyle Behaviours Be Improved in People with Familial Hypercholesterolemia? Results of a Parallel Randomised Controlled Trial. 2012.

283. Cox M, O'Connor C, Biggs K, Hind D, Bortolami O, Franklin M, et al. The feasibility of early pulmonary rehabilitation and activity after COPD exacerbations: External pilot randomised controlled trial, qualitative case study and exploratory economic evaluation. Health Technol Assess (Rockv). 2018;22:1–204.

284. DeVito Dabbs A, Song MKK, Hawkins R, Aubrecht J, Kovach K, Terhorst L, et al. An intervention fidelity framework for technology-based behavioral interventions. Nurs Res. 2011;60:340–7.

285. Di Rezze B, Law M, Eva K, Pollock N, Gorter JW. The Paediatric Rehabilitation

Observational Measure of Fidelity (PROF): Development and validation. 2013.

286. Di Rezze B, Law M, Eva K, Pollock N, Gorter JW. Therapy behaviours in paediatric rehabilitation: essential attributes for intervention with children with physical disabilities. Disabil Rehabil. 2014;36:16–22.

287. Salamh PA, Liu X, Hanney WJ, Sprague PA, Kolber MJ, Baron JS, et al. Promoting psychosocial well-being following stroke: study protocol for a randomized, controlled trial. BMJ Open. 2018;9:1173–86.

288. Dillon L, Clemson L, Coxon K, Keay L. Understanding the implementation and efficacy of a home-based strength and balance fall prevention intervention in people aged 50 years or over with vision impairment: a process evaluation protocol. BMC Health Serv Res. 2018;18:512.

289. Fortier CB, Kenna A, Dams-O'Connor K, Fonda J, Levin LK, Hursh C, et al. Feasibility of a Skills-Based Group Reintegration Workshop for OEF/OIF Veterans: STEP-Home. 2018.

290. French CT, Diekemper RL, Irwin RS, Adams TM, Altman KW, Barker AF, et al. Assessment of Intervention Fidelity and Recommendations for Researchers Conducting Studies on the Diagnosis and Treatment of Chronic Cough in the Adult: CHEST Guideline and Expert Panel Report. Chest. 2015;148:32–54.

291. Galaviz KI, Harden SM, Smith E, Blackman KC, Berrey LM, Mama SK, et al. Physical activity promotion in Latin American populations: a systematic review on issues of internal and external validity. Int J Behav Nutr Phys Act. 2014;11:77.

292. Gladwell PW, Cramp F, Palmer S. Matching the perceived benefits of Transcutaneous

Electrical Nerve Stimulation (TENS) for chronic musculoskeletal pain against Patient Reported Outcome Measures using the International Classification of Functioning, Disability and Health (ICF). 2019.

293. Kladouchou V, Papathanasiou I, Efstratiadou EA, Christaki V, Hilari K. Treatment integrity of elaborated semantic feature analysis aphasia therapy delivered in individual and group settings. Int J Lang Commun Disord. 2017;52:733–49.

294. Leeuw M, Goossens MEJB, de Vet HCW, Vlaeyen JWS. The fidelity of treatment delivery can be assessed in treatment outcome studies: a successful illustration from behavioral medicine. 2009.

295. Arbesman M, Lieberman D, Berlanstein DR. Method for the Evidence-Based Reviews on Occupational Therapy and Stroke. Am J Occup Ther. 2015;69:1–5.

296. McLennon SM, Hancock RD, Redelman K, Scarton LJ, Riley E, Sweeney B, et al. Comparing treatment fidelity between study arms of a randomized controlled clinical trial for stroke family caregivers. 2016.

297. Moore SA, Hrisos N, Flynn D, Errington L, Price C, Avery L. How should long-term free-living physical activity be targeted after stroke? A systematic review and narrative synthesis. Int J Behav Nutr Phys Act. 2018;15:100.

298. Sandborgh M, Lindberg P, Åsenlöf P, Denison E. Implementing behavioural medicine in physiotherapy treatment. Part I: clinical trial. Adv Physiother. 2010;12:2–12.

299. Sandborgh M, Åsenlöf P, Lindberg P, Denison E. Implementing behavioural medicine in physiotherapy treatment. Part II: Adherence to treatment protocol. 2010.

300. Skidmore ER, Dawson DR, Whyte EM, Butters MA, Dew MA, Grattan ES, et al. Developing complex interventions: lessons learned from a pilot study examining strategy training in acute stroke rehabilitation. Clin Rehabil. 2014;28:378–87.

301. Wilbur J, Schoeny ME, Buchholz SW, Fogg L, Miller AM, Braun LT, et al. Women's Lifestyle Physical Activity Program for African American Women: Fidelity Plan and Outcomes. J Phys Act Health. 2016;13:1100–9.

302. Woolf C, Caute A, Haigh Z, Galliers J, Wilson S, Kessie A, et al. A comparison of remote therapy, face to face therapy and an attention control intervention for people with aphasia: a quasi-randomised controlled feasibility study. Clin Rehabil. 2016;30:359–73.

303. Barber SE, Jackson C, Hewitt C, Ainsworth HR, Buckley H, Akhtar S, et al. Assessing the feasibility of evaluating and delivering a physical activity intervention for pre-school children: a pilot randomised controlled trial. Pilot feasibility Stud. 2016;2:12.

304. White P, Gilworth G, Lewin S, Hogg L, Tuffnell R, Taylor SJCC, et al. Improving uptake and completion of pulmonary rehabilitation in COPD with lay health workers: feasibility of a clinical trial. Int J Chron Obstruct Pulmon Dis. 2019;

305. Robbins-Welty GA, Mueser L, Mitchell C, Pope N, Arnold R, Park SY, et al. Interventionist training and intervention fidelity monitoring and maintenance for CONNECT, a nurse-led primary palliative care in oncology trial. Contemp Clin Trials Commun. 2018;10:57–61.

306. McDowell K, O'Neill B, Blackwood B, Clarke C, Gardner E, Johnston P, et al. Effectiveness of an exercise programme on physical function in patients discharged from hospital following critical illness: A randomised controlled trial (the REVIVE trial). Thorax.

2016;72:600–9.

307. Broekhuizen K, Jelsma GM J, van PoppelNM M, Koppes LJ L, Brug J, van Mechelen W. Is the process of delivery of an individually tailored lifestyle intervention associated with improvements in LDL cholesterol and multiple lifestyle behaviours in people with Familial Hypercholesterolemia? BMC Public Health. 2012;12:348.

308. Broekhuizen K, van Poppel MNM, Koppes LLJ, Brug J, van Mechelen W. A tailored lifestyle intervention to reduce the cardiovascular disease risk of individuals with Familial Hypercholesterolemia (FH): design of the PRO-FIT randomised controlled trial. 2010.

309. Yates BC, Schumacher KL, Norman JE, Krogstrand KS, Meza J, Shurmur S. Intervention fidelity in a translational study: lessons learned. Res Theory Nurs Pract. 2013;27:131–48.

310. Robins JL, Jallo N, Kinser PA. Treatment Fidelity in Mind–Body Interventions. J Holist Nurs. 2019;37:189–99.

311. Palsola M, Renko E, Kostamo K, Lorencatto F, Hankonen N. Thematic analysis of acceptability and fidelity of engagement for behaviour change interventions: The Let's Move It intervention interview study. Br J Health Psychol. 2020;25:772–89.

312. Boden I, Skinner EH, Browning L, Reeve J, Anderson L, Hill C, et al. Preoperative physiotherapy for the prevention of respiratory complications after upper abdominal surgery: Pragmatic, double blinded, multicentre randomised controlled trial. BMJ. 2018;360.

313. Galea Holmes MN, Wileman V, Hassan S, Denning J, Critchley D, Norton S, et al. Physiotherapy informed by Acceptance and Commitment Therapy for chronic low back

pain: A mixed-methods treatment fidelity evaluation. Br J Health Psychol. 2022;27:935–55.

314. Clark F, Pyatak EA, Carlson M, Blanche EI, Vigen C, Hay J, et al. Implementing trials of complex interventions in community settings: the USC-Rancho Los Amigos pressure ulcer prevention study (PUPS). Clin Trials. 2014;11:218–29.

315. Stevens AB, Strasser DC, Uomoto J, Bowen SE, Falconer JA, AB S, et al. Utility of Treatment Implementation methods in clinical trial with rehabilitation teams. J Rehabil Res Dev. 2007;44:537–46.

316. Sekhon M, Cartwright M, Francis JJ. Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. BMC Health Serv Res. 2017;17:88.

317. Köykkä K, Absetz P, Araújo-Soares V, Knittle K, Sniehotta FF, Hankonen N. Combining the reasoned action approach and habit formation to reduce sitting time in classrooms: Outcome and process evaluation of the Let's Move It teacher intervention. J Exp Soc Psychol. 2019;81:27–38.

318. Hankonen N, Heino MTJ, Araujo-Soares V, Sniehotta FF, Sund R, Vasankari T, et al. 'Let's Move It' – a school-based multilevel intervention to increase physical activity and reduce sedentary behaviour among older adolescents in vocational secondary schools: a study protocol for a cluster-randomised trial. BMC Public Health. 2016;16:451.

319. Simmons RG, Walters ST, Pappas LM, Boucher KM, Boonyasiriwat W, Gammon A, et al. Implementation of best practices regarding treatment fidelity in the family colorectal

cancer awareness and risk education randomized controlled trial. SAGE Open. 2014;4.

320. Leventhal H, Friedman MA. Does Establishing Fidelity of Treatment Help in Understanding Treatment Efficacy? Comment on Bellg et al. (2004). Heal Psychol. 2004;23:452–6.

321. Wilson MG, Basta TB, Bynum BH, DeJoy DM, Vandenberg RJ, Dishman RK. Do intervention fidelity and dose influence outcomes? Results from the Move to Improve worksite physical activity program. Health Educ Res. 2010;25:294–305.

322. Poltawski L, Norris M, Dean S. Interventi on fidelity: Deve loping an experience -based model for rehabilitati on research. J Rehabil Med. 2014;46:609–15.

323. Allan LM, Wheatley A, Flynn E, Smith A, Fox C, Howel D, et al. Is it feasible to deliver a complex intervention to improve the outcome of falls in people with dementia? A protocol for the DIFRID feasibility study. Pilot feasibility Stud. 2018;4:170.

324. Söderlund A, Bring A, Åsenlöf P. A three-group study, internet-based, face-to-face based and standard- management after acute whiplash associated disorders (WAD) - Choosing the most efficient and cost-effective treatment: Study protocol of a randomized controlled trial. BMC Musculoskelet Disord. 2009;10.

325. Blanche EI, Fogelberg D, Diaz J, Carlson M, Clark F. Manualization of occupational therapy interventions: Illustrations from the pressure ulcer prevention research program. Am J Occup Ther. 2011;65:711–9.

326. Beneciuk JM, George SZ, Greco CM, Schneider MJ, Wegener ST, Saper RB, et al. Targeted interventions to prevent transitioning from acute to chronic low back pain in high-risk patients: development and delivery of a pragmatic training course of psychologically

informed physical therapy for the TARGET trial. Trials. 2019;20:256.

327. Åsenlöf P, Denison E, Lindberg P. Individually tailored treatment targeting activity, motor behavior, and cognition reduces pain-related disability: A randomized controlled trial in patients with musculoskeletal pain. 2005.

328. Crossley KM, Vicenzino B, Lentzos J, Schache AG, Pandy MG, Ozturk H, et al. Exercise, education, manual-therapy and taping compared to education for patellofemoral osteoarthritis: a blinded, randomised clinical trial. Osteoarthr Cartil. 2015;23:1457–64.

329. Vaughan-Graham J, Cott C, Wright F V. The Bobath/NDT concept in adult neurological rehabilitation: What is the state of the knowledge? A scoping review. 2015.

330. Van Stan JH, Dijkers MP, Whyte J, Hart T, Turkstra LS, Zanca JM, et al. The Rehabilitation Treatment Specification System: Implications for Improvements in Research Design, Reporting, Replication, and Synthesis. Arch Phys Med Rehabil. 2019;100:146–55.

331. Tougas ME, Hayden JA, McGrath PJ, Huguet A, Rozario S. A Systematic Review Exploring the Social Cognitive Theory of Self-Regulation as a Framework for Chronic Health Condition Interventions. Courvoisier DS, editor. PLoS One. 2015;10:e0134977.

332. Dijkers MP, Hart T, Tsaousides T, Whyte J, Zanca JM. Treatment taxonomy for rehabilitation: Past, present, and prospects. Arch Phys Med Rehabil. 2014;95:S6–16.

333. Whyte J. Contributions of Treatment Theory and Enablement Theory to Rehabilitation Research and Practice. Arch Phys Med Rehabil. 2014;95:S17-S23.e2.

334. Hart T. Treatment definition in complex rehabilitation interventions. Neuropsychol Rehabil. 2009;19:824–40.

335. Webb TL, Joseph J, Yardley L, Michie S. Using the Internet to Promote Health Behavior Change: A Systematic Review and Meta-analysis of the Impact of Theoretical Basis, Use of Behavior Change Techniques, and Mode of Delivery on Efficacy. J Med Internet Res. 2010;12:e4.

336. Glanz K, Bishop DB. The Role of Behavioral Science Theory in Development and Implementation of Public Health Interventions. Annu Rev Public Health. 2010;31:399–418.

337. Poltawski L, Briggs J, Forster A, Goodwin VA, James M, Taylor RS, et al. Informing the design of a randomised controlled trial of an exercise-based programme for long term stroke survivors: lessons from a before-and-after case series study. BMC Res Notes. 2013;6:324.

338. Blencowe N, Boddy A, Harris A, Hanna T, Whiting P, Cook J, et al. Accounting for intervention complexity in RCTs in surgery: New approaches for intervention definition and methods for monitoring fidelity. Trials Nov 29, 2013 p. O86.

339. Bickman L. The functions of program theory. New Dir Progr Eval. 1987;1987:5–18.

340. Agha RA, Borrelli MR, Farwana R, Koshy K, Fowler AJ, Orgill DP, et al. The SCARE 2018 statement: Updating consensus Surgical CAse REport (SCARE) guidelines. Int J Surg. 2018;60:132–6.

341. Agha RA, Borrelli MR, Farwana R, Koshy K, Fowler AJ, Orgill DP, et al. The PROCESS 2018 statement: Updating Consensus Preferred Reporting Of CasE Series in Surgery (PROCESS) guidelines. Int J Surg. 2018;60:279–82.

342. Toomey E, Matthews J, Guerin S, Hurley DA. Development of a Feasible Implementation

Within a Complex Physical Therapy – Led Self-Management Intervention. Phys Ther. 2016;96:1287–98.

343. LeLaurin J, Schmitzberger M, Eliazar-Macke N, Freytes IM, Dang S, Uphold C. A commentary on methodological issues in stroke caregiver research: lessons learned from three RESCUE intervention studies. 2019.

344. Birken M, Henderson C, Slade M. The development of an occupational therapy intervention for adults with a diagnosed psychotic disorder following discharge from hospital. Pilot Feasibility Stud. 2018;4.

345. O'Neill B, O'Shea O, McDonough S, McGarvey L, Bradbury I, Arden M, et al. Clinician-Facilitated Physical Activity Intervention Versus Pulmonary Rehabilitation for Improving Physical Activity in COPD: A Feasibility Study. COPD J Chronic Obstr Pulm Dis. 2018;15:254–64.

346. Schaaf RC, Benevides T, Mailloux Z, Faller P, Hunt J, Van Hooydonk E, et al. An intervention for sensory difficulties in children with autism: A randomized trial. 2014.

347. Lenker JA, Fuhrer MJ, Jutai JW, Demers L, Scherer MJ, DeRuyter F. Treatment theory, intervention specification, and treatment fidelity in assistive technology outcomes research. Assist Technol. 2010;22:129–40.

348. Sheill A, Hawe P, Gold L. Complex interventions or complex systems? Implications for health economic evaluation. Bmj. 2008;336:1281–3.

349. Beard DJ, Davies L, Cook JA, Stokes J, Leal J, Fletcher H, et al. Rehabilitation versus surgical reconstruction for non-acute anterior cruciate ligament injury (ACL SNNAP): a pragmatic

randomised controlled trial. Lancet. 2022;400:605–15.

350. Ownsworth T, Fleming J, Tate R, Beadle E, Griffin J, Kendall M, et al. Do People With Severe Traumatic Brain Injury Benefit From Making Errors? A Randomized Controlled Trial of Error-Based and Errorless Learning. Neurorehabil Neural Repair. 2017;31:1072–82.

351. Holt RI, Hind D, Gossage-Worrall R, Bradburn MJ, Saxon D, McCrone P, et al. Structured lifestyle education to support weight loss for people with schizophrenia, schizoaffective disorder and first episode psychosis: the STEPWISE RCT. Health Technol Assess. 2018;22:1–160.

352. Morrison JD, Becker H, Stuifbergen AK. Evaluation of Intervention Fidelity in a Multisite Clinical Trial in Persons With Multiple Sclerosis. J Neurosci Nurs. 2017;49:344–8.

353. Thomas LH, Watkins CL, French B, Sutton C, Forshaw D, Cheater F, et al. Study protocol: ICONS: identifying continence options after stroke: a randomised trial. Trials. 2011;12:131.

354. Malden S, Hughes AR, Gibson AM, Bardid F, Androutsos O, De Craemer M, et al. Adapting the ToyBox obesity prevention intervention for use in Scottish preschools: Protocol for a feasibility cluster randomised controlled trial. 2018.

355. Jørgensen MB, Faber A, Jespersen T, Hansen K, Ektor-Andersen J, Hansen J V, et al. Implementation of physical coordination training and cognitive behavioural training interventions at cleaning workplaces--secondary analyses of a randomised controlled trial. 2012.

356. Logan A, Freeman J, Kent B, Pooler J, Creanor S, Vickery J, et al. Standing Practice In

Rehabilitation Early after Stroke (SPIRES): a functional standing frame programme (prolonged standing and repeated sit to stand) to improve function and quality of life and reduce neuromuscular impairment in people with severe sub-acu. Pilot feasibility Stud. 2018;4:66.

357. Åsenlöf P, Denison E, Lindberg P. Long-term follow-up of tailored behavioural treatment and exercise based physical therapy in persistent musculoskeletal pain: A randomized controlled trial in primary care. Eur J Pain. 2009;13:1080–8.

358. Roberts JL, Pritchard AW, Williams M, Totton N, Morrison V, Din NU, et al. Mixed methods process evaluation of an enhanced community-based rehabilitation intervention for elderly patients with hip fracture. BMJ Open. 2018;8:e021486.

359. Williamson E, Ward L, Vadher K, Dutton SJ, Parker B, Petrou S, et al. Better Outcomes for Older people with Spinal Trouble (BOOST) Trial: A randomised controlled trial of a combined physical and psychological intervention for older adults with neurogenic claudication, a protocol. BMJ Open. 2018;8.

360. Masterson-Algar P, Burton CR, Rycroft-Malone J. The generation of consensus guidelines for carrying out process evaluations in rehabilitation research. BMC Med Res Methodol. 2018;18.

361. Wesson J, Clemson L, Brodaty H, Lord S, Taylor M, Gitlin L, et al. A feasibility study and pilot randomised trial of a tailored prevention program to reduce falls in older people with mild dementia. BMC Geriatr. 2013;13:89.

362. Baron JS, Sullivan KJ, Swaine JM, Aspinall A, Jaglal S, Presseau J, et al. Self-management

interventions for skin care in people with a spinal cord injury: part 2-a systematic review of use of theory and quality of intervention reporting. Spinal Cord. 2018;56:837–46.

363. Stein KF, Sargent JT, Rafaels N. Establishing Fidelity of the Independent Variable in Nursing Clinical Trials. Nurs Res. 2007;56:54–62.

364. Bissonnette JM. Adherence: a concept analysis. J Adv Nurs. 2008;63:634–43.

365. Argent R, Daly A, Caulfield B. Patient Involvement With Home-Based Exercise Programs: Can Connected Health Interventions Influence Adherence? JMIR mHealth uHealth. 2018;6:e47.

366. Gladwell PWW illia. WW illia., Badlan K, Cramp F, Palmer S. Problems, Solutions, and Strategies Reported by Users of Transcutaneous Electrical Nerve Stimulation for Chronic Musculoskeletal Pain: Qualitative Exploration Using Patient Interviews. Phys Ther. 2016;96:1039–48.

367. Barber SE, Akhtar S, Jackson C, Bingham DD, Hewitt C, Routen A, et al. Preschoolers in the Playground: a pilot cluster randomised controlled trial of a physical activity intervention for children aged 18 months to 4 years. Public Heal Res. 2015;3:1–210.

368. Harwood RH, van der Wardt V, Goldberg SE, Kearney F, Logan P, Hood-Moore V, et al. A development study and randomised feasibility trial of a tailored intervention to improve activity and reduce falls in older adults with mild cognitive impairment and mild dementia. Pilot feasibility Stud. 2018;4:49.

369. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for Implementation Research: Conceptual Distinctions, Measurement Challenges, and

Research Agenda. Adm Policy Ment Heal Ment Heal Serv Res. 2011;38:65–76.

370. Pyatak EA, Carandang K, Vigen C, Blanchard J, Sequeira PA, Wood JR, et al. Resilient, Empowered, Active Living with Diabetes (REAL Diabetes) study: Methodology and baseline characteristics of a randomized controlled trial evaluating an occupation-based diabetes management intervention for young adults. Contemp Clin Trials. 2017;54:8–17.

371. Burkart S, St Laurent CW, Alhassan S. Process evaluation of a culturally-tailored physical activity intervention in African-American mother-daughter dyads. Prev Med reports. 2017;8:88–92.

372. Tuntland H, Aaslund MK, Espehaug B, Førland O, Kjeken I. Reablement in community-dwelling older adults: A randomised controlled trial Physical functioning, physical health and activity. BMC Geriatr. 2015;15.

373. Sturkenboom IH, Nijhuis-van der Sanden MW, Graff MJ. A process evaluation of a home-based occupational therapy intervention for Parkinson's patients and their caregivers performed alongside a randomized controlled trial. Clin Rehabil. 2016;30:1186–99.

374. Godfrey E, Galea Holmes M, Wileman V, Mccracken L, Norton S, Moss-Morris R, et al. Physiotherapy informed by Acceptance and Commitment Therapy (PACT): Protocol for a randomised controlled trial of PACT versus usual physiotherapy care for adults with chronic low back pain. BMJ Open. 2016;6.

375. Sweeney-Magee M, Kale D, Galton S, Hamill A, Gilbert H. Assessing the fidelity of delivery of an intervention to increase attendance at the English Stop Smoking Services. Implement Sci. 2016;11:166.

376. James K, Quirk A, Patterson S, Brennan G, Stewart D. Quality of intervention delivery in a cluster randomised controlled trial: A qualitative observational study with lessons for fidelity. Trials. 2017;18:1–10.

377. Cousins SE, Blencowe NS, Tsang C, Chalmers KA, Mardanpour A, Carr A, et al. Optimising the design of invasive placebo interventions in randomised controlled trials. Br J Surg. 2019;in press.

378. Probst P, Grummich K, Harnoss JC, Hüttner FJ, Jensen K, Braun S, et al. Placebo-Controlled Trials in Surgery. Medicine (Baltimore). 2016;95:e3516.

379. Tracey I. Getting the pain you expect: Mechanisms of placebo, nocebo and reappraisal effects in humans. Nat Med. 2010;16:1277–83.

380. Maddocks M, Kerry R, Turner A, Howick J. Problematic placebos in physical therapy trials. J Eval Clin Pract. 2016;22:598–602.

381. Niemansburg SL, van Delden JJM, Dhert WJA, Bredenoord AL. Reconsidering the ethics of sham interventions in an era of emerging technologies. Surgery. 2015;157:801–10.

382. Fregni F, Imamura M, Chien HF, Lew HL, Boggio P, Kaptchuk TJ, et al. Challenges and Recommendations for Placebo Controls in Randomized Trials in Physical and Rehabilitation Medicine. Am J Phys Med Rehabil. 2010;89:160–72.

383. Bialosky JE, Bishop MD, Cleland JA. Individual Expectation: An Overlooked, but Pertinent, Factor in the Treatment of Individuals Experiencing Musculoskeletal Pain. Phys Ther. 2010/06/30. 2010;90:1345–55.

384. Campaner R, Cerri M. Manipulative evidence and medical interventions: some

qualifications. Hist Philos Life Sci. 2020;42:15.

385. Burchett HED, Blanchard L, Kneale D, Thomas J. Assessing the applicability of public health intervention evaluations from one setting to another: a methodological study of the usability and usefulness of assessment tools and frameworks. Heal Res Policy Syst. 2018;16:88.

386. Ziemann A, Brown L, Sadler E, Ocloo J, Boaz A, Sandall J. Influence of external contextual factors on the implementation of health and social care interventions into practice within or across countries—a protocol for a 'best fit' framework synthesis. Syst Rev. 2019;8:258.

387. Greaves CJ, Sheppard KE, Abraham C, Hardeman W, Roden M, Evans PH, et al. Systematic review of reviews of intervention components associated with increased effectiveness in dietary and physical activity interventions. BMC Public Health. 2011;11:119.

388. Schroter S, Glasziou P, Heneghan C. Quality of descriptions of treatments: A review of published randomised controlled trials. BMJ Open. 2012;2:1–7.

389. Glasziou P, Meats E, Heneghan C, Shepperd S. What is missing from descriptions of treatment in trials and reviews? BMJ. 2008;336:1472–4.

390. Rothwell PM. External validity of randomised controlled trials: "To whom do the results of this trial apply?" Lancet. 2005;365:82–93.

391. Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. BMJ. 2015;350:h2147–h2147.

392. Candy B, Vickerstaff V, Jones L, King M. Description of complex interventions: Analysis of changes in reporting in randomised trials since 2002. Trials. 2018;19:1–9.

393. Breitenstein SM, Gross D, Garvey CA, Hill C, Fogg L, Resnick B. Implementation fidelity in community-based interventions. Res Nurs Health. 2010;33:164–73.

394. Denis JL, Hebert Y, Langley A, Lozeau D, Trottier LH. Explanation Diffusion Patterns of Complex Health Care Innovations. Heal Care Manag Rev. 2002;27:60–73.

395. McCreight MS, Rabin BA, Glasgow RE, Ayele RA, Leonard CA, Gilmartin HM, et al. Using the Practical, Robust Implementation and Sustainability Model (PRISM) to qualitatively assess multilevel contextual factors to help plan, implement, evaluate, and disseminate health services programs. Transl Behav Med. 2019;9:1002–11.

396. Ribeiro DC, Spiers K, Thomas L, Leilua K, Wilkes M, Norton S, et al. Monitoring, implementation and reporting of interventions in a selection of trials assessing exercise therapy for the management of shoulder subacromial pain: a cross-sectional investigation. BMJ Open. 2021;11:e044462.

397. Johnson-Kozlow M, Hovell MF, Rovniak LS, Sirikulvadhana L, Wahlgren DR, Zakarian JM. Fidelity issues in secondhand smoking interventions for children. Nicotine Tob Res. 2008;10:1677–90.

398. Hampton LH, Kaiser AP. Intervention effects on spoken-language outcomes for children with autism: a systematic review and meta-analysis. J Intellect Disabil Res. 2016;60:444–63.

399. Moher D, Schulz KF, Simera I, Altman DG. Guidance for Developers of Health Research Reporting Guidelines. PLoS Med. 2010;7:e1000217.

400. Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, et al. Guidelines for

Reporting Reliability and Agreement Studies (GRRAS) were proposed. J Clin Epidemiol. 2011;64:96–106.

401. National Archives ot the United States of America. Title 45 - Public Welfare Subtitle A - Department of Health and Human Services Subchapter A - General Administration Part 46 - Protection of Human Subjects Subpart A - Basic HHS Policy for Protection of Human Research Subjects. Code Fed Regul. 2023;104:1–4.

402. Smith-Lock KM, Leitao S, Lambert L, Nickels L. Effective intervention for expressive grammar in children with specific language impairment. Int J Lang Commun Disord. 2013;48:265–82.

403. Brosseau L, Wells GA, Kenny GP, Reid R, Maetzel A, Tugwell P, et al. The implementation of a community-based aerobic walking program for mild to moderate knee osteoarthritis: A knowledge translation randomized controlled trial: Part II: Clinical outcomes. BMC Public Health. 2012;12.

404. Chaipinyo K, Karoonsupcharoen O. No difference between home-based strength training and home-based balance training on pain in patients with knee osteoarthritis: a randomised trial. Aust J Physiother. 2009;55:25–30.

405. Manske RC, Meschke M, Porter A, Smith B, Reiman M. A Randomized Controlled Single-Blinded Comparison of Stretching Versus Stretching and Joint Mobilization for Posterior Shoulder Tightness Measured by Internal Rotation Motion Loss. Sport Heal A Multidiscip Approach. 2010;2:94–100.

406. Boyce LK, Innocenti MS, Roggman LA, Norman VKJ, Ortiz E. Telling Stories and Making

Books: Evidence for an Intervention to Help Parents in Migrant Head Start Families Support Their Children's Language and Literacy. Early Educ Dev. 2010;21:343–71.

407. Cadmus L, Patrick MB, MacIejewski ML, Topolski T, Belza B, Patrick DL. Community-based aquatic exercise and quality of life in persons with osteoarthritis. Med Sci Sports Exerc. 2010;42:8–15.

408. Ağlamış B, Toraman NF, Yaman H. The effect of a 12-week supervised multicomponent exercise program on knee OA in Turkish women. J Back Musculoskelet Rehabil. 2008;21:121–8.

409. Huebner CE. Promoting Toddlers' Language Development Through Community-Based Intervention. J Appl Dev Psychol. 2000;21:513–35.

410. Dwyer L, Parkin-Smith GF, Brantingham JW, Korporaal C, Cassa TK, Globe G, et al. Manual and Manipulative Therapy in Addition to Rehabilitation for Osteoarthritis of the Knee: Assessor-Blind Randomized Pilot Trial. J Manipulative Physiol Ther. 2015;38:1-21.e2.

411. Jassil FC, Manning S, Lewis N, Steinmo S, Kingett H, Lough F, et al. Feasibility and Impact of a Combined Supervised Exercise and Nutritional-Behavioral Intervention following Bariatric Surgery: A Pilot Study. J Obes. 2015;2015:1–12.

412. Cheung NW, Cinnadaio N, Russo M, Marek S. A pilot randomised controlled trial of resistance exercise bands in the management of sedentary subjects with type 2 diabetes. Diabetes Res Clin Pract. 2009;83:e68–71.

413. Bennell KL, Kyriakides M, Metcalf B, Egerton T, Wrigley T V., Hodges PW, et al. Neuromuscular versus quadriceps strengthening exercise in patients with medial knee

osteoarthritis and varus malalignment: A randomized controlled trial. Arthritis Rheumatol. 2014;66:950–9.

414. Balducci S, Zanuso S, Nicolucci A, De Feo P, Cavallo S, Cardelli P, et al. Effect of an intensive exercise intervention strategy on modifiable cardiovascular risk factors in subjects with type 2 diabetes mellitus: a randomized controlled trial: the Italian Diabetes and Exercise Study (IDES). Arch Intern Med. 2010;170:1794–803.

415. Gallagher AL, Chiat S. Evaluation of speech and language therapy interventions for pre-school children with specific language impairment: a comparison of outcomes following specialist intensive, nursery-based and no intervention. Int J Lang Commun Disord. 2009;44:616–38.

416. Plotnikoff RC, Eves N, Jung M, Sigal RJ, Padwal R, Karunamuni N. Multicomponent, home-based resistance training for obese adults with type 2 diabetes: a randomized controlled trial. Int J Obes. 2010;34:1733–41.

417. Hardan AY, Gengoux GW, Berquist KL, Libove RA, Ardel CM, Phillips J, et al. A randomized controlled trial of Pivotal Response Treatment Group for parents of children with autism. J Child Psychol Psychiatry. 2015;56:884–92.

418. Lonigan CJ, Whitehurst GJ. Relative efficacy of parent and teacher involvement in a shared-reading intervention for preschool children from low-income backgrounds. Early Child Res Q. 1998;13:263–90.

419. Stensvold D, Viken H, Steinshamn SL, Dalen H, Støylen A, Loennechen JP, et al. Effect of exercise training for five years on all cause mortality in older adults—the Generation 100

study: randomised controlled trial. BMJ. 2020;371:m3485.

420. Deyle GD, Allison SC, Matekel RL, Ryder MG, Stang JM, Gohdes DD, et al. Physical Therapy Treatment Effectiveness for Osteoarthritis of the Knee: A Randomized Comparison of Supervised Clinical Exercise and Manual Therapy Procedures Versus a Home Exercise Program. Phys Ther. 2005;85:1301 LP – 1317.

421. Solomon R, Van Egeren LA, Mahoney G, Quon Huber MS, Zimmerman P. PLAY Project Home Consultation Intervention Program for Young Children With Autism Spectrum Disorders. J Dev Behav Pediatr. 2014;35:475–85.

422. Siller M, Hutman T, Sigman M. A Parent-Mediated Intervention to Increase Responsive Parental Behaviors and Child Communication in Children with ASD: A Randomized Clinical Trial. J Autism Dev Disord. 2013;43:540–55.

423. Pile E, Girolametto L, Johnson CJ, Chen X, Cleave PL. Shared Book Reading Intervention for Children with Language Impairment: Using Parents-as-aides in Language Intervention. Can J Speech-Language Pathol Audiol. 2010;34:96–109.

424. Tyler AA, Gillon G, Macrae T, Johnson RL. Direct and Indirect Effects of Stimulating Phoneme Awareness vs. Other Linguistic Skills in Preschoolers With Co-occurring Speech and Language Impairments. Top Lang Disord. 2011;31:128–44.

425. Pollard H, Ward G, Hoskins W, Hardy K. The effect of a manual therapy knee protocol on osteoarthritic knee pain: a randomised controlled trial. J Can Chiropr Assoc. 2008;52:229–42.

426. Bonett DG. Sample Size Requirements for Testing and Estimating Coefficient Alpha. J Educ

Behav Stat. 2002;27:335–40.

427. Manis KT, Choi D. The virtual reality hardware acceptance model (VR-HAM): Extending and individuating the technology acceptance model (TAM) for virtual reality hardware. J Bus Res. 2019;100:503–13.

428. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika. 1951;16:297–334.

429. Cohen J. A Coefficient of Agreement for Nominal Scales. Educ Psychol Meas. 1960;20:37–46.

430. Mokkink LB, Terwee CB, Gibbons E, Stratford PW, Alonso J, Patrick DL, et al. Inter-rater agreement and reliability of the COSMIN (COnsensus-based Standards for the selection of health status Measurement Instruments) Checklist. BMC Med Res Methodol. 2010;10:82.

431. Lemons DS. An Introduction to Stochastic Processes in Physics. Baltimore, Md.: John Hopkins University Press; 2002. 34–36 p.

432. Ljunggren G, Perera NKP, Hägglund M. Inter-rater Reliability in Assessing Exercise Fidelity for the Injury Prevention Exercise Programme Knee Control in Youth Football Players. Sport Med - Open. 2019;5:35.

433. O'Brien J, Finch CF. The Implementation of Musculoskeletal Injury-Prevention Exercise Programmes in Team Ball Sports: A Systematic Review Employing the RE-AIM Framework. Sport Med. 2014;44:1305–18.

434. Taber KS. The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. Res Sci Educ. 2018;48:1273–96.

435. Dechartres A, Charles P, Hopewell S, Ravaud P, Altman DG. Reviews assessing the quality or the reporting of randomized controlled trials are increasing over time but raised questions about how quality is assessed. J Clin Epidemiol. 2011;64:136–44.

436. Giraudeau B, Higgins JPT, Tavernier E, Trinquart L. Sample size calculation for meta-epidemiological studies. Stat Med. 2016;35:239–50.

437. Stockard J. An analysis of the fidelity implementation policies of the what works clearinghouse. Curr Issues Educ. 2010;13:1–24.

438. Schemitsch EH, Bhandari M, Boden SD, Bourne RB, Bozic KJ, Jacobs JJ, et al. The evidence-based approach in bringing new orthopaedic devices to market. J Bone Jt Surg - Ser A. 2010;92:1030–7.

439. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011;343:1–9.

440. Guyatt G. Evidence-Based Medicine. JAMA. 1992;268:2420.

441. van Tulder MW, Suttorp M, Morton S, Bouter LM, Shekelle P. Empirical Evidence of an Association Between Internal Validity and Effect Size in Randomized Controlled Trials of Low-Back Pain. Spine (Phila Pa 1976). 2009;34:1685–92.

442. Berlin JA, Golub RM. Meta-analysis as Evidence. JAMA. 2014;312:603.

443. Page MJ. Controversy and Debate on Meta-epidemiology. Paper 4: Confounding and other concerns in meta-epidemiological studies of bias. J Clin Epidemiol. 2020;123:133–4.

444. Herbert RD. Controversy and Debate on Meta-epidemiology. Paper 2: Meta-epidemiological studies of bias may themselves be biased. J Clin Epidemiol.

2020;123:127–30.

445. Hempel S, Miles J, Booth M, Wang Z, Johnsen B, Morton S, et al. Detection of Associations Between Trial Quality and Effect Sizes. 2012;

446. Welton NJ, Ades AE, Carlin JB, Altman DG, Sterne JAC. Models for potentially biased evidence in meta-analysis using empirically based priors. J R Stat Soc Ser A (Statistics Soc. 2009;172:119–36.

447. Schulz KF, Chalmers I, Hayes R, Altman DG. Empirical Evidence of Bias: Dimensions of Methodological Quality Associated With Estimates of Treatment Effects in Controlled Trials. JAMA. 1995;273:408.

448. Dechartres A, Trinquart L, Faber T, Ravaud P. Empirical evaluation of which trial characteristics are associated with treatment effect estimates. J Clin Epidemiol. 2016;77:24–37.

449. Hempel S, Suttorp MJ, Miles JN V., Wang Z, Maglione M, Mroton S, et al. Empirical Evidence of Associations Between Trial Quality and Effect Size Methods Research Report Empirical Evidence of Associations Between Trial Quality and Effect Size. Rockville, Maryland; 2012.

450. Murad MH, Wang Z. Guidelines for reporting meta-epidemiological methodology research. Evid Based Med. 2017;22:139–42.

451. Trinquart L, Diard E, Centre of Research in Epidemiology and Statistics Sorbonne Paris Cité. Power and sample size calculation for meta-epidemiological studies [Internet]. Power and sample size calculation for meta-epidemiological studies. Paris: Centre of Research in

Epidemiology and Statistics Sorbonne Paris Cité (CRESS UMR1153); 2015 [cited 2020 Jan 3]. Available from: http://www.clinicalepidemio.fr/metaepidemio/

452. Tanaka TL. Estimating the impact of assessment and treatment fidelity on aphasia treatment outcomes. University of New Mexico; 2017.

453. Lee E, Dobbins M, Decorby K, McRae L, Tirilis D, Husson H. An optimal search filter for retrieving systematic reviews and meta-analyses. BMC Med Res Methodol. 2012;12.

454. StataCorp LLC. Stata Statistical Software: Release 16. College Station, TX: StataCorp LLC; 2019.

455. Coppock A. randomizr: Easy-to-Use Tools for Common Forms of Random Assignment and Sampling. 2021.

456. Akiba CF, Powell BJ, Pence BW, Nguyen MXB, Golin C, Go V. The case for prioritizing implementation strategy fidelity measurement: benefits and challenges. Transl Behav Med. 2022;12:335–42.

457. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd ed. New York: Routledge; 2013.

458. DerSimonian R, Kacker R. Random-effects model for meta-analysis of clinical trials: An update. Contemp Clin Trials. 2007;28:105–14.

459. Cohen J. Chapter 1: the concepts of power analysis. In: Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale, NJ: Academic Press, Inc; 1988. p. 1–17.

460. Ryan R. Heterogeneity and subgroup analyses in Cochrane Consumers and Communication Review Group reviews: Planning the analysis at protocol stage. Cochrane

Consum Commun Rev Gr. 2014;2014:2–9.

461. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst. 1959;22:719–48.

462. Savović J, Turner RM, Mawdsley D, Jones HE, Beynon R, Higgins JPT, et al. Association Between Risk-of-Bias Assessments and Results of Randomized Trials in Cochrane Reviews: The ROBES Meta-Epidemiologic Study. Am J Epidemiol. 2018;187:1113–22.

463. Nuesch E, Trelle S, Reichenbach S, Rutjes AWS, Tschannen B, Altman DG, et al. Small study effects in meta-analyses of osteoarthritis trials: meta-epidemiological study. BMJ. 2010;341:c3515–c3515.

464. Avery L, Flynn D, van Wersch A, Sniehotta FF, Trenell MI. Changing Physical Activity Behavior in Type 2 Diabetes. Diabetes Care. 2012;35:2681–9.

465. Bellicha A, Ciangura C, Poitou C, Portero P, Oppert JM. Effectiveness of exercise training after bariatric surgery-a systematic literature review and meta-analysis. Obes Rev. 2018;19:1544–56.

466. Briani RV, Ferreira AS, Pazzinatto MF, Pappas E, De Oliveira Silva D, Azevedo FM de. What interventions can improve quality of life or psychosocial factors of individuals with knee osteoarthritis? A systematic review with meta-analysis of primary outcomes from randomised controlled trials. Br J Sports Med. 2018;52:1031–8.

467. Finch M, Jones J, Yoong S, Wiggers J, Wolfenden L. Effectiveness of centre-based childcare interventions in increasing child physical activity: a systematic review and meta-analysis for policymakers and practitioners. Obes Rev. 2016;17:412–28.

468. McMichan L, Gibson AM, Rowe DA. Classroom-based physical activity and sedentary behavior interventions in adolescents: A systematic review and meta-analysis. J Phys Act Heal. 2018;15:383–93.

469. Young HML, March DS, Graham-Brown MPM, Jones AW, Curtis F, Grantham CS, et al. Effects of intradialytic cycling exercise on exercise capacity, quality of life, physical function and cardiovascular measures in adult haemodialysis patients: a systematic review and meta-analysis. Nephrol Dial Transplant. 2018;33:1436–45.

470. Maynard BR, Peters KE, Vaughn MG, Sarteschi CM. Fidelity in After-School Program Intervention Research: A Systematic Review. Res Soc Work Pract. 2013;23:613–23.

471. Naleppa MJ, Cagle JG. Treatment Fidelity in Social Work Intervention Research: A Review of Published Studies. Res Soc Work Pract. 2010;20:674–81.

472. Herbert RD, B⃠ K. Analysis of quality of interventions in systematic reviews. BMJ. 2005;331:507–9.

473. Miller FL, O'Connor DP, Herring MP, Sailors MH, Jackson AS, Dishman RK, et al. Exercise Dose, Exercise Adherence, and Associated Health Outcomes in the TIGER Study. Med Sci Sport Exerc. 2014;46:69–75.

474. Taylor KL, Weston M, Batterham AM. Evaluating Intervention Fidelity: An Example from a High-Intensity Interval Training Study. Piacentini MF, editor. PLoS One. 2015;10:e0125166.

475. Sawilowsky SS. New Effect Size Rules of Thumb. J Mod Appl Stat Methods. 2009;8:597–9.

476. Powell R, McGregor G, Ennis S, Kimani PK, Underwood M. Is exercise-based cardiac rehabilitation effective? A systematic review and meta-analysis to re-examine the

evidence. BMJ Open. 2018;8:e019656.

477. Yoon IA, Slade K, Fazel S. Outcomes of psychological therapies for prisoners with mental health problems: A systematic review and meta-analysis. J Consult Clin Psychol. 2017;85:783–802.

478. Miller WR, Rollnick S. The effectiveness and ineffectiveness of complex behavioral interventions: Impact of treatment fidelity. Contemp Clin Trials. 2014;37:234–41.

479. Klein RA, Vianello M, Hasselman F, Adams BG, Adams RB, Alper S, et al. Many Labs 2: Investigating Variation in Replicability Across Samples and Settings. Adv Methods Pract Psychol Sci. 2018;1:443–90.

480. Honglertnapakul W, Cavuoto KM, McKeown CA, Capó H. Surgical treatment of strabismus in thyroid eye disease: characteristics, dose–response, and outcomes. J Am Assoc Pediatr Ophthalmol Strabismus. 2020;24:72.e1-72.e7.

481. Scherrenberg M, Janssen J, Lauwers A, Schreurs I, Swinnen M, Dendale P. Is there an optimal dose of cardiac rehabilitation in coronary artery disease patients? Int J Cardiol. 2021;330:7–11.

482. Ludyga S, Gerber M, Pühse U, Looser VN, Kamijo K. Systematic review and meta-analysis investigating moderators of long-term effects of exercise on cognition in healthy individuals. Nat Hum Behav. 2020;4:603–12.

483. Bolvig J, Juhl CB, Boutron I, Tugwell P, Ghogomu EAT, Pardo JP, et al. Some Cochrane risk-of-bias items are not important in osteoarthritis trials: a meta-epidemiological study based on Cochrane reviews. J Clin Epidemiol. 2018;95:128–36.

484. Kjaergard LL, Villumsen J, Gluud C. Reported Methodologic Quality and Discrepancies between Large and Small Randomized Trials in Meta-Analyses. Ann Intern Med. 2001;135:982.

485. Kjaergard LL, Nikolova D, Gluud C. Randomized clinical trials inHepatology: Predictors of quality. Hepatology. 1999;30:1134–8.

486. Hariohm K, Jeyanthi S, Kumar JS, Prakash V. Description of interventions is under-reported in physical therapy clinical trials. Brazilian J Phys Ther. 2017;21:281–6.

487. Olivo SA, Macedo LG, Gadotti IC, Fuentes J, Stanton T, Magee DJ. Scales to Assess the Quality of Randomized Controlled Trials: A Systematic Review. Phys Ther. 2008;88:156–75.

488. Levine SZ, Leucht S. Identifying a system of predominant negative symptoms: Network analysis of three randomized clinical trials. Schizophr Res. 2016;178:17–22.

489. Vale CL, Tierney JF, Burdett S. Can trial quality be reliably assessed from published reports of cancer trials: evaluation of risk of bias assessments in systematic reviews. BMJ. 2013;346:f1798–f1798.

490. Mhaskar R, Djulbegovic B, Magazin A, Soares HP, Kumar A. Published methodological quality of randomized controlled trials does not reflect the actual quality assessed in protocols. J Clin Epidemiol. 2012;65:602–9.

491. Austin PC, Steyerberg EW. The number of subjects per variable required in linear regression analyses. J Clin Epidemiol. 2015;68:627–36.

492. Geissbühler M, Hincapié CA, Aghlmandi S, Zwahlen M, Jüni P, da Costa BR. Most published

meta-regression analyses based on aggregate data suffer from methodological pitfalls: a meta-epidemiological study. BMC Med Res Methodol. 2021;21:123.

493. Kirkham JJ, Dwan KM, Altman DG, Gamble C, Dodd S, Smyth R, et al. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. BMJ. 2010;340:c365–c365.

494. Devereaux PJ, Choi PTL, El-Dika S, Bhandari M, Montori VM, Schünemann HJ, et al. An observational study found that authors of randomized controlled trials frequently use concealment of randomization and blinding, despite the failure to report these methods. J Clin Epidemiol. 2004;57:1232–6.

495. Gamerman V, Cai T, Elsäßer A. Pragmatic randomized clinical trials: best practices and statistical guidance. Heal Serv Outcomes Res Methodol. 2019;19:23–35.

496. Olson MB, McCreedy EM, Baier RR, Shield RR, Zediker EE, Uth R, et al. Measuring implementation fidelity in a cluster-randomized pragmatic trial: development and use of a quantitative multi-component approach. Trials. 2022;23:43.

497. Perry M, Drašković I, Lucassen P, Vernooij-Dassen M, van Achterberg T, Rikkert MO. Effects of educational interventions on primary dementia care: A systematic review. Int J Geriatr Psychiatry. 2011;26:1–11.

498. Hohenschurz-Schmidt D, Kleykamp BA, Draper-Rodi J, Vollert J, Chan J, Ferguson M, et al. Pragmatic trials of pain therapies: a systematic review of methods. Pain. 2022;163:21–46.

499. Washington T, Zimmerman S, Cagle J, Reed D, Cohen L, Beeber AS, et al. Fidelity Decision Making in Social and Behavioral Research: Alternative Measures of Dose and Other

Considerations. Soc Work Res. 2014;38:154–62.

500. Cahill LS, Carey LM, Lannin NA, Turville M, Neilson CL, Lynch EA, et al. Implementation interventions to promote the uptake of evidence-based practices in stroke rehabilitation. Cochrane Database Syst Rev. 2020;10:CD012575.

501. Palmer JA, Parker VA, Barre LR, Mor V, Volandes AE, Belanger E, et al. Understanding implementation fidelity in a pragmatic randomized clinical trial in the nursing home setting:a mixed-methods examination. Trials. 2019;20:1–10.

502. Blaschke TF, Osterberg L, Vrijens B, Urquhart J. Adherence to Medications: Insights Arising from Studies on the Unreliable Link Between Prescribed and Actual Drug Dosing Histories. Annu Rev Pharmacol Toxicol. 2012;52:275–301.

503. Pawson R. Pragmatic trials and implementation science: grounds for divorce? BMC Med Res Methodol. 2019;19:176.

504. Davies L, Cook J, Leal J, Areia CM, Shirkey B, Jackson W, et al. Comparison of the clinical and cost effectiveness of two management strategies (rehabilitation versus surgical reconstruction) for non-acute anterior cruciate ligament (ACL) injury: study protocol for the ACL SNNAP randomised controlled trial. Trials. 2020;21:405.

505. Chambers DA, Glasgow RE, Stange KC. The dynamic sustainability framework: addressing the paradox of sustainment amid ongoing change. Implement Sci. 2013;8:117.

506. Esmail LC, Barasky R, Mittman BS, Hickam DH. Improving Comparative Effectiveness Research of Complex Health Interventions: Standards from the Patient-Centered Outcomes Research Institute (PCORI). J Gen Intern Med. 2020;35:875–81.

507. Fitzpatrick C, Gillies C, Seidu S, Kar D, Ioannidou E, Davies MJ, et al. Effect of pragmatic versus explanatory interventions on medication adherence in people with cardiometabolic conditions: a systematic review and meta-analysis. BMJ Open. 2020;10:e036575.

508. Tuzzio L, Larson EB, Chambers DA, Coronado GD, Curtis LH, Weber WJ, et al. Pragmatic clinical trials offer unique opportunities for disseminating, implementing, and sustaining evidence-based practices into clinical care: Proceedings of a workshop. Healthcare. 2019;7:51–7.

509. Proctor EK, Landsverk J, Aarons G, Chambers D, Glisson C, Mittman B. Implementation Research in Mental Health Services: an Emerging Science with Conceptual, Methodological, and Training challenges. Adm Policy Ment Heal Ment Heal Serv Res. 2009;36:24–34.

510. Zimmerman S, Wretman CJ, Ward K, Tandan M, Sloane PD, Preisser JS. Fidelity and sustainability of Mouth Care Without a Battle and lessons for other innovations in care. Geriatr Nurs (Minneap). 2020;41:878–84.

511. Apter AJ. Understanding Adherence Requires Pragmatic Trials. JAMA Pediatr. 2015;169:310.

512. Musahl V, Karlsson J. Anterior Cruciate Ligament Tear. Solomon CG, editor. N Engl J Med. 2019;380:2341–8.

513. Sanders TL, Maradit Kremers H, Bryan AJ, Larson DR, Dahm DL, Levy BA, et al. Incidence of Anterior Cruciate Ligament Tears and Reconstruction. Am J Sports Med. 2016;44:1502–7.

514. Gabr A, De Medici A, Haddad F. The National Ligament Registry: The Fifth Annual Report

(2019). 2019;5.

515. Ajuied A, Wong F, Smith C, Norris M, Earnshaw P, Back D, et al. Anterior cruciate ligament injury and radiologic progression of knee osteoarthritis: a systematic review and meta-analysis. Am J Sports Med. 2014;42:2242–52.

516. Peck J. Long-term sequelae and management following anterior cruciate ligament injury. BMJ Case Rep. 2014;2014:bcr2014204239–bcr2014204239.

517. Monk AP, Hopewell S, Harris K, Davies LJ, Beard D, Price A. Surgical versus conservative interventions for treating anterior cruciate ligament injuries. In: Monk AP, editor. Cochrane Database of Systematic Reviews. Chichester, UK: John Wiley & Sons, Ltd; 2014.

518. Linko E, Harilainen A, Malmivaara A, Seitsalo S. Surgical versus conservative interventions for anterior cruciate ligament ruptures in adults. Linko E, editor. Cochrane Database Syst Rev. 2016;2016:CD001356.

519. Frobell RB, Roos HP, Roos EM, Roemer FW, Ranstam J, Lohmander LS. Treatment for acute anterior cruciate ligament tear: five year outcome of randomised trial. BMJ. 2013;346:f232–f232.

520. Bollen SR, Scott BW. Rupture of the anterior cruciate ligament — a quiet epidemic? Injury. 1996;27:407–9.

521. British Association for Surgery of he Knee, British Orthopaedic Sports Trauma and Arthroscopy Association. Best practice for management of anterior cruciate ligament (ACL) injuries. [Internet]. [cited 2022 Jun 3]. Available from: https://www.boa.ac.uk/static/88a4c3e3-df3e-4e51-a92e7d2f86d7d82a/Best-Practice-

Book-for-management-of-Anterior-Cruciate-Ligament-injuries.pdf

522. Stokes JR, Beard DJ, Davies L, Shirkey BA, Price A, Cook JA, et al. ACL Surgery Necessity in Non-Acute Patients (ACL SNNAP): a statistical analysis plan for a randomised controlled trial. Trials. 2022;23:389.

523. Paez A, Hind D, Mcculloch P, Beard D, Paez A, Hind D, et al. An integrated treatment fidelity framework for clinical trials of complex interventions in healthcare : a methodological systematic review and " best-fit " framework synthesis. PROSPERO: International prospective register of systematic reviews. 2019.

524. Andrade R, Pereira R, van Cingel R, Staal JB, Espregueira-Mendes J. How should clinicians rehabilitate patients after ACL reconstruction? A systematic review of clinical practice guidelines (CPGs) with a focus on quality appraisal (AGREE II). Br J Sports Med. 2020;54:512–9.

525. Tessier S. From Field Notes, to Transcripts, to Tape Recordings: Evolution or Combination? Int J Qual Methods. 2012;11:446–60.

526. Beard DJ, Davies L, Cook J, Stokes J, Leal J, Fletcher H, et al. The ACL SNNAP Trial: Anterior Cruciate Ligament Surgery Necessity in Non Acute Patients Long: Comparison of the clinical and cost effectiveness of two management strategies for non-acute Anterior Cruciate Ligament (ACL) injury: Rehabilitation vers. Health Technol Assess (Rockv). 2023;in press.

527. Fann JR, Hart T, Ciol MA, Moore M, Bogner J, Corrigan JD, et al. Improving transition from inpatient rehabilitation following traumatic brain injury: Protocol for the BRITE pragmatic

comparative effectiveness trial. Contemp Clin Trials. 2021;104:106332.

528. Teo JL, Zheng Z, Bird SR. Identifying the factors affecting 'patient engagement' in exercise rehabilitation. BMC Sports Sci Med Rehabil. 2022;14:18.

529. Essery R, Geraghty AWA, Kirby S, Yardley L. Predictors of adherence to home-based physical therapies: a systematic review. Disabil Rehabil. 2017;39:519–34.

530. Burns D, Boyer P, Razmjou H, Richards R, Whyne C. Adherence Patterns and Dose Response of Physiotherapy for Rotator Cuff Pathology: Longitudinal Cohort Study. JMIR Rehabil Assist Technol. 2021;8:e21374.

531. Gupta S. Intention-to-treat concept: A review. Perspect Clin Res. 2011;2:109.

532. Wiltsey Stirman S, Baumann AA, Miller CJ. The FRAME: an expanded framework for reporting adaptations and modifications to evidence-based interventions. Implement Sci. 2019;14:58.

533. Shelton RC, Cooper BR, Stirman SW. The Sustainability of Evidence-Based Interventions and Practices in Public Health and Health Care. Annu Rev Public Health. 2018;39:55–76.

534. Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. J Chronic Dis. 1967;20:637–48.

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

536. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: A revised tool for assessing risk of bias in randomised trials. BMJ. 2019;366:1–8.

537. McLean SM, Burton M, Bradley L, Littlewood C. Interventions for enhancing adherence with physiotherapy: A systematic review. Man Ther. 2010;15:514–21.

538. Cohen DJ, Crabtree BF, Etz RS, Balasubramanian BA, Donahue KE, Leviton LC, et al. Fidelity Versus Flexibility. Translating Evidence-Based Research into Practice. Am J Prev Med. 2008;35:381–9.

539. Yamato TP, Maher CG, Saragiotto BT, Catley MJ, Moseley AM. Rasch analysis suggested that items from the template for intervention description and replication (TIDieR) checklist can be summed to create a score. J Clin Epidemiol. 2018;101:28–34.

540. Hoffmann TC, Walker MF. 'TIDieR-ing up' the reporting of interventions in stroke research: The importance of knowing what is in the 'black box'. Int J Stroke. 2015;10:657–8.

541. Lindquist R, Wyman JF, Talley KMC, Findorff MJ, Gross CR. Design of Control-Group Conditions in Clinical Trials of Behavioral Interventions. J Nurs Scholarsh. 2007;39:214–21.

542. Hulleman CS, Cordray DS. Moving From the Lab to the Field: The Role of Fidelity and Achieved Relative Intervention Strength. J Res Educ Eff. 2009;2:88–110.

**Appendix I:** Chapter 1 supplementary materials

**Figure 1.1:** The CERT Checklist[151]

# CERT ✓ *C*onsensus on *E*xercise *R*eporting *T*emplate

## A Checklist for what to include when reporting exercise programs

| Section/Topic | Item # | Checklist item | Location ** | |
|---|---|---|---|---|
| | | | Primary paper (page, table, appendix) | † Other (paper or protocol, website (URL) |
| **WHAT: materials** | 1 | Detailed description of the type of exercise equipment (e.g. weights, exercise equipment such as machines, treadmill, bicycle ergometer etc) | | |
| **WHO: provider** | 2 | Detailed description of the qualifications, teaching/supervising expertise, and/or training undertaken by the exercise instructor | | |
| **HOW: delivery** | 3 | Describe whether exercises are performed individually or in a group | | |
| | 4 | Describe whether exercises are supervised or unsupervised and how they are delivered | | |
| | 5 | Detailed description of how adherence to exercise is measured and reported | | |
| | 6 | Detailed description of motivation strategies | | |
| | 7a | Detailed description of the decision rule(s) for determining exercise progression | | |
| | 7b | Detailed description of how the exercise program was progressed | | |
| | 8 | Detailed description of each exercise to enable replication (e.g. photographs, illustrations , video etc) | | |
| | 9 | Detailed description of any home program component (e.g. other exercises, stretching etc) | | |
| | 10 | Describe whether there are any non-exercise components (e.g. education, cognitive behavioural therapy, massage etc) | | |
| | 11 | Describe the type and number of adverse events that occurred during exercise | | |
| **WHERE: location** | 12 | Describe the setting in which the exercises are performed | | |
| **WHEN, HOW MUCH: dosage** | 13 | Detailed description of the exercise intervention including, but not limited to, number of exercise repetitions/sets/sessions, session duration, intervention/program duration etc | | |
| **TAILORING: what, how** | 14a | Describe whether the exercises are generic (one size fits all) or tailored whether tailored to the individual | | |
| | 14b | Detailed description of how exercises are tailored to the individual | | |
| | 15 | Describe the decision rule for determining the starting level at which people commence an exercise program (such as beginner, intermediate, advanced etc) | | |
| **HOW WELL: planned, actual** | 16a | Describe how adherence or fidelity to the exercise intervention is assessed/measured | | |
| | 16b | Describe the extent to which the intervention was delivered as planned | | |

***It is recommended that this checklist is used in conjunction with the Explanation and Elaboration Statement which is a guide each item in the CERT Checklist**

The CERT Checklist is designed for reporting details of an exercise intervention. The CERT Checklist should be used in conjunction with a reporting checklist appropriate for the study type e.g. the CONSORT Statement (www.consort-statement.org) for randomised controlled trials, the SPIRIT Statement (www.spirit-statement.org) for a clinical trial protocol. For further guidance regarding reporting guidelines please consult the EQUATOR network (www.equator-network.org)

** Authors – please use N/A if an item is not applicable  Reviewers – please use "?" if information is not provided or not/insufficiently reported

† If the information is not provided in the primary paper that is under consideration, please provide details of where this information is available e.g. in a published protocol, published papers (provide citation details) or on a website (provide the URL).

Figure 1.2: The SPIRIT Statement[26]


SPIRIT 2013 Checklist: Recommended items to address in a clinical trial protocol and related documents*

| Section/item | Item No | Description |
|---|---|---|
| **Administrative information** | | |
| Title | 1 | Descriptive title identifying the study design, population, interventions, and, if applicable, trial acronym |
| Trial registration | 2a | Trial identifier and registry name. If not yet registered, name of intended registry |
| | 2b | All items from the World Health Organization Trial Registration Data Set |
| Protocol version | 3 | Date and version identifier |
| Funding | 4 | Sources and types of financial, material, and other support |
| Roles and responsibilities | 5a | Names, affiliations, and roles of protocol contributors |
| | 5b | Name and contact information for the trial sponsor |
| | 5c | Role of study sponsor and funders, if any, in study design; collection, management, analysis, and interpretation of data; writing of the report; and the decision to submit the report for publication, including whether they will have ultimate authority over any of these activities |
| | 5d | Composition, roles, and responsibilities of the coordinating centre, steering committee, endpoint adjudication committee, data management team, and other individuals or groups overseeing the trial, if applicable (see Item 21a for data monitoring committee) |
| **Introduction** | | |
| Background and rationale | 6a | Description of research question and justification for undertaking the trial, including summary of relevant studies (published and unpublished) examining benefits and harms for each intervention |
| | 6b | Explanation for choice of comparators |
| Objectives | 7 | Specific objectives or hypotheses |
| Trial design | 8 | Description of trial design including type of trial (eg, parallel group, crossover, factorial, single group), allocation ratio, and framework (eg, superiority, equivalence, noninferiority, exploratory) |

**Methods: Participants, interventions, and outcomes**

| | | |
|---|---|---|
| Study setting | 9 | Description of study settings (eg, community clinic, academic hospital) and list of countries where data will be collected. Reference to where list of study sites can be obtained |
| Eligibility criteria | 10 | Inclusion and exclusion criteria for participants. If applicable, eligibility criteria for study centres and individuals who will perform the interventions (eg, surgeons, psychotherapists) |
| Interventions | 11a | Interventions for each group with sufficient detail to allow replication, including how and when they will be administered |
| | 11b | Criteria for discontinuing or modifying allocated interventions for a given trial participant (eg, drug dose change in response to harms, participant request, or improving/worsening disease) |
| | 11c | Strategies to improve adherence to intervention protocols, and any procedures for monitoring adherence (eg, drug tablet return, laboratory tests) |
| | 11d | Relevant concomitant care and interventions that are permitted or prohibited during the trial |
| Outcomes | 12 | Primary, secondary, and other outcomes, including the specific measurement variable (eg, systolic blood pressure), analysis metric (eg, change from baseline, final value, time to event), method of aggregation (eg, median, proportion), and time point for each outcome. Explanation of the clinical relevance of chosen efficacy and harm outcomes is strongly recommended |
| Participant timeline | 13 | Time schedule of enrolment, interventions (including any run-ins and washouts), assessments, and visits for participants. A schematic diagram is highly recommended (see Figure) |
| Sample size | 14 | Estimated number of participants needed to achieve study objectives and how it was determined, including clinical and statistical assumptions supporting any sample size calculations |
| Recruitment | 15 | Strategies for achieving adequate participant enrolment to reach target sample size |

**Methods: Assignment of interventions (for controlled trials)**

Allocation:

| | | |
|---|---|---|
| Sequence generation | 16a | Method of generating the allocation sequence (eg, computer-generated random numbers), and list of any factors for stratification. To reduce predictability of a random sequence, details of any planned restriction (eg, blocking) should be provided in a separate document that is unavailable to those who enrol participants or assign interventions |

| | | |
|---|---|---|
| Allocation concealment mechanism | 16b | Mechanism of implementing the allocation sequence (eg, central telephone; sequentially numbered, opaque, sealed envelopes), describing any steps to conceal the sequence until interventions are assigned |
| Implementation | 16c | Who will generate the allocation sequence, who will enrol participants, and who will assign participants to interventions |
| Blinding (masking) | 17a | Who will be blinded after assignment to interventions (eg, trial participants, care providers, outcome assessors, data analysts), and how |
| | 17b | If blinded, circumstances under which unblinding is permissible, and procedure for revealing a participant's allocated intervention during the trial |
| **Methods: Data collection, management, and analysis** | | |
| Data collection methods | 18a | Plans for assessment and collection of outcome, baseline, and other trial data, including any related processes to promote data quality (eg, duplicate measurements, training of assessors) and a description of study instruments (eg, questionnaires, laboratory tests) along with their reliability and validity, if known. Reference to where data collection forms can be found, if not in the protocol |
| | 18b | Plans to promote participant retention and complete follow-up, including list of any outcome data to be collected for participants who discontinue or deviate from intervention protocols |
| Data management | 19 | Plans for data entry, coding, security, and storage, including any related processes to promote data quality (eg, double data entry; range checks for data values). Reference to where details of data management procedures can be found, if not in the protocol |
| Statistical methods | 20a | Statistical methods for analysing primary and secondary outcomes. Reference to where other details of the statistical analysis plan can be found, if not in the protocol |
| | 20b | Methods for any additional analyses (eg, subgroup and adjusted analyses) |
| | 20c | Definition of analysis population relating to protocol non-adherence (eg, as randomised analysis), and any statistical methods to handle missing data (eg, multiple imputation) |
| **Methods: Monitoring** | | |
| Data monitoring | 21a | Composition of data monitoring committee (DMC); summary of its role and reporting structure; statement of whether it is independent from the sponsor and competing interests; and reference to where further details about its charter can be found, if not in the protocol. Alternatively, an explanation of why a DMC is not needed |

| | | |
|---|---|---|
| | 21b | Description of any interim analyses and stopping guidelines, including who will have access to these interim results and make the final decision to terminate the trial |
| Harms | 22 | Plans for collecting, assessing, reporting, and managing solicited and spontaneously reported adverse events and other unintended effects of trial interventions or trial conduct |
| Auditing | 23 | Frequency and procedures for auditing trial conduct, if any, and whether the process will be independent from investigators and the sponsor |

**Ethics and dissemination**

| | | |
|---|---|---|
| Research ethics approval | 24 | Plans for seeking research ethics committee/institutional review board (REC/IRB) approval |
| Protocol amendments | 25 | Plans for communicating important protocol modifications (eg, changes to eligibility criteria, outcomes, analyses) to relevant parties (eg, investigators, REC/IRBs, trial participants, trial registries, journals, regulators) |
| Consent or assent | 26a | Who will obtain informed consent or assent from potential trial participants or authorised surrogates, and how (see Item 32) |
| | 26b | Additional consent provisions for collection and use of participant data and biological specimens in ancillary studies, if applicable |
| Confidentiality | 27 | How personal information about potential and enrolled participants will be collected, shared, and maintained in order to protect confidentiality before, during, and after the trial |
| Declaration of interests | 28 | Financial and other competing interests for principal investigators for the overall trial and each study site |
| Access to data | 29 | Statement of who will have access to the final trial dataset, and disclosure of contractual agreements that limit such access for investigators |
| Ancillary and post-trial care | 30 | Provisions, if any, for ancillary and post-trial care, and for compensation to those who suffer harm from trial participation |
| Dissemination policy | 31a | Plans for investigators and sponsor to communicate trial results to participants, healthcare professionals, the public, and other relevant groups (eg, via publication, reporting in results databases, or other data sharing arrangements), including any publication restrictions |
| | 31b | Authorship eligibility guidelines and any intended use of professional writers |
| | 31c | Plans, if any, for granting public access to the full protocol, participant-level dataset, and statistical code |

Figure 1.3 The TIDieR Checklist[150]

**The TIDieR (Template for Intervention Description and Replication) Checklist*:**

Information to include when describing an intervention and the location of the information

| Item number | Item | Where located ** | |
|---|---|---|---|
| | | Primary paper (page or appendix number) | Other † (details) |
| | **BRIEF NAME** | | |
| 1. | Provide the name or a phrase that describes the intervention. | _____ | _____ |
| | **WHY** | | |
| 2. | Describe any rationale, theory, or goal of the elements essential to the intervention. | _____ | _____ |
| | **WHAT** | | |
| 3. | Materials: Describe any physical or informational materials used in the intervention, including those provided to participants or used in intervention delivery or in training of intervention providers. Provide information on where the materials can be accessed (e.g. online appendix, URL). | _____ | _____ |
| 4. | Procedures: Describe each of the procedures, activities, and/or processes used in the intervention, including any enabling or support activities. | _____ | _____ |
| | **WHO PROVIDED** | | |
| 5. | For each category of intervention provider (e.g. psychologist, nursing assistant), describe their expertise, background and any specific training given. | _____ | _____ |
| | **HOW** | | |
| 6. | Describe the modes of delivery (e.g. face-to-face or by some other mechanism, such as internet or telephone) of the intervention and whether it was provided individually or in a group. | _____ | _____ |
| | **WHERE** | | |
| 7. | Describe the type(s) of location(s) where the intervention occurred, including any necessary infrastructure or relevant features. | _____ | _____ |
| | **WHEN and HOW MUCH** | | |
| 8. | Describe the number of times the intervention was delivered and over what period of time including the number of sessions, their schedule, and their duration, intensity or dose. | _____ | _____ |
| | **TAILORING** | | |
| 9. | If the intervention was planned to be personalised, titrated or adapted, then describe what, why, when, and how. | _____ | _____ |
| | **MODIFICATIONS** | | |
| 10.‡ | If the intervention was modified during the course of the study, describe the changes (what, why, when, and how). | _____ | _____ |
| | **HOW WELL** | | |
| 11. | Planned: If intervention adherence or fidelity was assessed, describe how and by whom, and if any strategies were used to maintain or improve fidelity, describe them. | _____ | _____ |
| 12.‡ | Actual: If intervention adherence or fidelity was assessed, describe the extent to which the intervention was delivered as planned. | _____ | _____ |

** **Authors** - use N/A if an item is not applicable for the intervention being described. **Reviewers** – use '?' if information about the element is not reported/not sufficiently reported.

† If the information is not provided in the primary paper, give details of where this information is available. This may include locations such as a published protoc or other published papers (provide citation details) or a website (provide the URL).

‡ If completing the TIDieR checklist for a protocol, these items are not relevant to the protocol and cannot be described until the study is complete.

* We strongly recommend using this checklist in conjunction with the TIDieR guide (see *BMJ* 2014;348:g1687) which contains an explanation and elaboration for each item.

* The focus of TIDieR is on reporting details of the intervention elements (and where relevant, comparison elements) of a study. Other elements and methodological features studies are covered by other reporting statements and checklists and have not been duplicated as part of the TIDieR checklist. When a **randomised trial** is being reported, th TIDieR checklist should be used in conjunction with the CONSORT statement (see www.consort-statement.org) as an extension of **Item 5 of the CONSORT 2010 Statement.** When a **clinical trial protocol** is being reported, the TIDieR checklist should be used in conjunction with the SPIRIT statement as an extension of **Item 11 of the SPIRIT 2013 Statement** (see www.spirit-statement.org). For alternate study designs, TIDieR can be used in conjunction with the appropriate checklist for that study design (see

**Appendix II:** Chapter II supplementary materials

Figure 2.1 Search strategies: Pubmed

((((((((((((((((((((("framework") OR "model") OR "procedure") OR "assessment") OR "process") OR "monitoring") OR "monitored") OR "strategy") OR "indicators")) AND ((((((((((("treatment integrity") OR "treatment fidelity") OR "treatment adherence") OR "treatment implementation") OR "treatment delivery") OR "treatment enactment") OR "intervention fidelity") OR "intervention integrity") OR "intervention adherence") OR "intervention implementation") OR "intervention delivery") OR "implementation strategy" AND ( "2006/01/01"[PDat] : "2019/12/31"[PDat] )))))) AND "physiotherapy")) OR ((((((((((((((("framework") OR "model") OR "procedure") OR "assessment") OR "process") OR "monitoring") OR "monitored") OR "strategy") OR "indicators")) AND ((((((((((("treatment integrity") OR "treatment fidelity") OR "treatment adherence") OR "treatment implementation") OR "treatment delivery") OR "treatment enactment") OR "intervention fidelity") OR "intervention integrity") OR "intervention adherence") OR "intervention implementation") OR "intervention delivery") OR "implementation strategy" AND ( "2006/01/01"[PDat] : "2019/12/31"[PDat] )))))) AND "physical therapy")) OR ((((((((((((((("framework") OR "model") OR "procedure") OR "assessment") OR "process") OR "monitoring") OR "monitored") OR "strategy") OR "indicators")) AND ((((((((((("treatment integrity") OR "treatment fidelity") OR "treatment adherence") OR "treatment implementation") OR "treatment delivery") OR "treatment enactment") OR "intervention fidelity") OR "intervention integrity") OR "intervention adherence") OR "intervention implementation") OR "intervention delivery") OR "implementation strategy" AND ( "2006/01/01"[PDat] : "2019/12/31"[PDat] )))))) AND "rehabilitation")) OR ((((((((((((((("framework") OR "model") OR "procedure") OR "assessment") OR "process") OR "monitoring") OR "monitored") OR "strategy") OR "indicators")) AND ((((((((((("treatment integrity") OR "treatment fidelity") OR "treatment adherence") OR "treatment implementation") OR "treatment delivery") OR "treatment enactment") OR "intervention fidelity") OR "intervention integrity") OR "intervention adherence") OR "intervention implementation") OR "intervention delivery") OR "implementation strategy" AND ( "2006/01/01"[PDat] : "2019/12/31"[PDat] )))))) AND "surgery")) OR ((((((((((((((("framework") OR "model") OR "procedure") OR "assessment") OR "process") OR "monitoring") OR "monitored") OR "strategy") OR "indicators")) AND ((((((((((("treatment integrity") OR "treatment fidelity") OR "treatment adherence") OR "treatment implementation") OR "treatment delivery") OR "treatment enactment") OR "intervention fidelity") OR "intervention integrity") OR "intervention adherence") OR "intervention implementation") OR "intervention delivery") OR "implementation strategy" AND ( "2006/01/01"[PDat] : "2019/12/31"[PDat] )))))) AND "occupational therapy")) OR ((((((((((((((("framework") OR "model") OR "procedure") OR "assessment") OR "process") OR "monitoring") OR "monitored") OR "strategy") OR "indicators")) AND ((((((((((("treatment integrity") OR "treatment fidelity") OR "treatment adherence") OR "treatment implementation") OR "treatment delivery") OR

"treatment enactment") OR "intervention fidelity") OR "intervention integrity") OR "intervention adherence") OR "intervention implementation") OR "intervention delivery") OR "implementation strategy" AND ( "2006/01/01"[PDat] : "2019/12/31"[PDat] )))))) AND speech therapy)) OR ((((((((((((((("framework") OR "model") OR "procedure") OR "assessment") OR "process") OR "monitoring") OR "monitored") OR "strategy") OR "indicators")) AND ((((((((((("treatment integrity") OR "treatment fidelity") OR "treatment adherence") OR "treatment implementation") OR "treatment delivery") OR "treatment enactment") OR "intervention fidelity") OR "intervention integrity") OR "intervention adherence") OR "intervention implementation") OR "intervention delivery") OR "implementation strategy" AND ( "2006/01/01"[PDat] : "2019/12/31"[PDat] )))))) AND physiatry)

Search strategies: SCOPUS

( ALL ( "treatment fidelity" OR "treatment integrity" OR "Treatment implementation" OR "treatment enactment" OR "intervention fidelity" OR "intervention integrity" OR "intervention implementation" OR "intervention enactment" OR “fidelity” OR “compliance” OR “concordance” or “adherence” OR “agreement” OR “delivery” OR “manualized” OR “treatment manual” OR “intervention manual” OR “per protocol” ) AND ALL (therapy OR process OR monitoring OR monitored OR framework OR assessment OR procedure OR strategy OR model OR indicator ) AND TITLE-ABS-KEY ( "physical therapy" OR physiotherapy OR surgery OR physiatry OR "occupational therapy" OR "speech therapy" OR rehabilitation ) ) AND PUBYEAR > 2004 AND ( EXCLUDE ( SUBJAREA , "PSYC" ) OR EXCLUDE ( SUBJAREA , "SOCI" ) OR EXCLUDE ( SUBJAREA , "BIOC" ) OR EXCLUDE ( SUBJAREA , "ARTS" ) OR EXCLUDE ( SUBJAREA , "PHAR" ) OR EXCLUDE ( SUBJAREA , "ENGI" ) OR EXCLUDE ( SUBJAREA , "AGRI" ) OR EXCLUDE ( SUBJAREA , "ENVI" ) OR EXCLUDE ( SUBJAREA , "PHYS" ) OR EXCLUDE ( SUBJAREA , "BUSI" ) OR EXCLUDE ( SUBJAREA , "CENG" ) OR EXCLUDE ( SUBJAREA , "DENT" ) OR EXCLUDE ( SUBJAREA , "EART" ) OR EXCLUDE ( SUBJAREA , "IMMU" ) OR EXCLUDE ( SUBJAREA , "ECON" ) )

Figure 2.1 Search strategies: **Embase**

| Set# | Searched for | Results |
|---|---|---|
| S39 | (S30 OR S31 OR S32 OR S33 OR S36 OR S37) and (pd(20050101-20191205)) | 721° |
| S38 | S30 OR S31 OR S32 OR S33 OR S36 OR S37 | 758° |
| S37 | S27 AND physiotherapy | 89° |
| S36 | S27 AND (speech therapy) | 46° |
| S35 | S27 AND ("speech therapy") | 11° |
| S34 | S27 AND physiatry | 0° |
| S33 | S27 AND ("occupational therapy") | 46° |
| S32 | S27 AND surgery | 399° |
| S31 | S27 AND rehabilitation | 278° |
| S30 | S29 OR S28 | 212° |
| S29 | S27 AND (physical therapy) | 212° |
| S28 | S27 AND ("physical therapy") | 51° |
| S27 | S26 AND S5 | 2437° |
| S26 | S13 OR S25 | 3733° |
| S25 | ("concordance") OR ("treatment implementation") OR (“treatment adherence”) OR (“treatment compliance”) OR | 1362° |
| S13 | ("intervention fidelity") OR ("intervention adherence") OR ("intervention implementation") OR ("intervention integrity") OR ("intervention delivery") OR (“intervention adherence”) | 2446° |
| S12 | ("treatment fidelity") OR ("treatment adherence") OR ("treatment implementation") OR ("treatment integrity") | 10297* |
| S5 | framework OR model OR process OR procedure OR strategy OR assessment OR indicat$ OR monitor$ OR checklist | 9301276 |

* Duplicates are removed from the search but included in the result count.

° Duplicates are removed from the search and from the result count.

Results: 721, 166 “Hits” after Screening

Figure 2.1 Search strategies: CINHAL

| # | Query | Last Run Via | Results |
|---|---|---|---|
| S11 | (S4 OR S5 OR S6 OR S7) AND (S9) | Interface - EBSCOhost Research Databases<br>Search Screen - Advanced Search<br>Database - CINAHL Complete | 245 |
| S10 | S4 OR S5 OR S6 OR S7 | Interface - EBSCOhost Research Databases<br>Search Screen - Advanced Search<br>Database - CINAHL Complete | 706,700 |
| S9 | (S3) AND (S8) | Interface - EBSCOhost Research Databases<br>Search Screen - Advanced Search<br>Database - CINAHL Complete | 1,487 |
| S8 | ( framework or model or theory) OR (process) OR (procedure ) OR (strategy) | Interface - EBSCOhost Research Databases Search Screen - Advanced Search<br>Database - CINAHL Complete | 1,091,682 |
| S7 | physical medicine and rehabilitation | Interface - EBSCOhost Research Databases Search Screen - Advanced Search<br>Database - CINAHL Complete | 5,152 |
| S6 | complex interventions in health | Interface - EBSCOhost Research Databases Search Screen - Advanced Search | 15 |
| S5 | surgery or operation or surgical procedure or surgical treatment | Interface - EBSCOhost Research Databases Search Screen - Advanced Search<br>Database - CINAHL Complete | 492,957 |
| S4 | ( physiotherapy or physical therapy or rehabilitation ) OR occupational therapy OR ( speech therapy or speech pathology or speech language pathology ) | Interface - EBSCOhost Research Databases Search Screen - Advanced Search<br>Database - CINAHL Complete | 232,880 |
| S3 | (S1) OR (S2) | Interface - EBSCOhost Research Databases Search Screen - Advanced Search<br>Database - CINAHL Complete | 3,136 |
| S2 | treatment delivery OR intervention delivery | Interface - EBSCOhost Research Databases Search Screen - Advanced Search<br>Database - CINAHL Complete | 1,074 |
| S1 | TX ( treatment integrity or treatment fidelity ) OR intervention integrity OR intervention fidelity | Interface - EBSCOhost Research Databases Search Screen - Advanced Search<br>Database - CINAHL Complete | 2,120 |

Table 2.1: Papers included in the Best-fit framework synthesis

| Authors | Year | Field/ Discpl. | Study design | Fidelity monitoring, measurement, or support method | Fid. Model | Fidelity Model Components, if reported , or direct quote/passage about fidelity extracted from paper. |
|---|---|---|---|---|---|---|
| Abell, et al | 2015 | CR PA | SR | TIDIER items 10-12 | NR | When and How Much Describes the dose/schedule of the intervention including the following: (a) Intensity The intensity of exercise used in the intervention (e.g., % heart rate) (b) Frequency The frequency of exercise sessions (c) Session Time The duration of each individual exercise session (d) Overall Duration The overall duration of the exercise intervention 9 Tailoring Describes the what, why, when, and how of intervention titration, personalization, or progression 10 Modifications Describes any modifications to the intervention during the course of the study 11 How well: planned Describes strategies used to maintain or improve fidelity (how and by whom) 12 How well: actual Describes the extent to which the intervention was delivered as planned (if adherence or fidelity was assessed) |
| Adams, et al | 2012 | SLP | RCT | Audit of planned intervention sessions versus received sessions and adherence to written activity procedure as stated in the manual | Reported elsewhere | Full details in Adams, C. and Gaile, J., Forthcoming 2012, Managing Children's Pragmatic and Social Communication Needs in the Early School Years ( Manchester : Roundway Centre Publ.) (available at: http://www.roundwaycentre.org.uk/publications). |
| Allan, et al | 2018 | OT, PT, RH | PTCL FS | Assessment of fidelity to components (protocol?) | NR | NR |
| Allan, et al | 2018 | OT, PT, RH | PTCL FS | Assessment of fidelity to components (protocol?) | NR | NR |
| Anderson, et al | 2012 | PT PA | RCT | Inferred from proxy. | NR | NR |
| Ang, et al | 2012 | PA | PTCL | Audio taped contacts to provide ongoing feedback and coaching to the MI health practitioner, on-going supervision and feedback. 25% of the audio taped phone contacts were formally evaluated for treatment integrity using the Motivational Interviewing Treatment Integrity (MITI) method. The MITI evaluators were blind to randomization. | NR | NR |
| Ang,et al | 2011 | PA | RCT | Phone contacts were taped to provide ongoing feedback and coaching to the therapists. supervision and | NR | NR |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | feedback and checklist for integrity. | | |
| Asenlof , et al | 2009 | PT | RCT | Patients' reports on treatment content, PT session by session documentation of treatment content, as well as individual working sheets were collected in order to obtain an estimate of the proportions of participants in each condition that received the program as was intended. | Bellg (NIH-BCC) | NR<br><br>Fidelity: Treatment dose was equivalent and within the stipulated range within and across conditions. Maintenance of treatment provider skills. |
| Asenlof , et al | 2005 | PT | RCT | Patient reports of treatment content, therapists' documentation of treatment content session by session, and individual working sheets were collected to obtain an estimate of the proportions of subjects in each condition that received the program as it was intended. | Bellg NIH-BCC | NR<br><br>Fidelity: intervention was implemented with all the planned components. treatment dose monitored. |
| Aunger , et al | 2019 | PA | PTCL (RCT FS) | self-reported checklist, document/ recording assessment checklist. | Fidelity to theory. | NR |
| Avery, et al | 2014 | PA LI | PTCL RCT | standardised training, study manuals, ongoing process evaluation and video recording and assessment of patient sessions (intervention and usual clinical care) with fidelity check-list developed for the study | NR<br><br>Cites Bellg NIH-BCC | NR<br><br>Fidelity: presence/absence and appropriate use of intervention components. fidelity of delivery (adherence to intervention components). Adherence to the study protocol (fidelity assessment |
| Barber et al | 2015 | PA | FS | Assessed with NIH-BCC guidance. observer attended sessions in the initiation phase at each intervention school; intervention delivery was scored on a scale from 1 to 4. | NIH-BCC | Fidelity: Adherence to the intervention protocol, Delivery as per manual. |
| Barber et al | 2016 | PA | PILOT RCT | assessed according to NIH Behaviour Change Consortium guidance | NIH-BCC | Fidelity: Delivery as per manual |

| | | | | | | |
|---|---|---|---|---|---|---|
| Baron et al | 2016 | RH | SR | TIDieR used to assess reporting of fidelity and adherence | Bellg/NIH-BCC | Fidelity of intervention delivery is extremely important as efficacy can only be determined if an intervention has been delivered as intended. |
| Bavan, et al | 2019 | PT PA | FS US | n/a | | |
| Beneciuk, et al | 2019 | PT | PT | n/a | Borelli NIH-BCC | Fidelity framework consisting of five domains (i.e., study design, training of providers, treatment delivery, treatment receipt, and treatment enactment. The PIPT program was designed to promote treatment fidelity by providing quality training that impacted key provider factors and that could be replicated. Thus, we incorporated quality improvement strategies (PIPT treatment checklist and booster training) and measures (physical therapist attitudes, beliefs, and confidence, described in greater detail below) to enhance treatment quality and the impact of training. |
| Bennet, et al | 2011 | Pain PT | SR | n/a | Carroll | This is defined as the degree to which the intervention and control are delivered and assessed as intended. |
| Bergström et al | 2013 | PA | cRCT | fidelity was measured as the dose delivered. | NR | Intervention fidelity, defined as the extent to which a programme adheres to its programme theory (Fraser, 2009), is crucial to understanding the causal mechanisms (Mercer, DeVinney, Fine, Green, & Dougherty, 2007). In this study fidelity was measured as the dose delivered. |
| Birken, et al | 2018 | OT | DS | standardised training on use of the intervention, and ensuring skill acquisition during the training through active discussion of the occupational therapists' role during GLOW. An intervention Fidelity checklist was developed to monitor the extent to which the clinicians adhered to the content of the intervention manual. | Bellg NIH-BCC | NR<br><br>NIH-BCC model used in Bellg." fidelity to the intervention was enhanced, as recommended by Bellg and colleagues [27] by providing standardised training on use of the intervention, and ensuring skill acquisition during the training through active discussion of the occupational therapists' role during GLOW." |
| Blanche, et al | 2011 | OT | PP | Manualisation of therapeutic intervention recommended, with fidelity check against manual. | NIH-BCC | NIH-BCC: study design, training of providers, delivery of treatment, receipt of treatment, and enactment of treatment skills…. development of a treatment manual that includes information about treatment dose (length and number of contacts) and the specific content of each contact, standardization of therapist training, monitoring of the intervention with fidelity checklists, and inclusion of strategies to measure the client's comprehension and enactment of the intervention principles addressed. |
| Blencowe et al | 2015 | Surg | PP | n/a | Carroll 2007 | During a trial, it may be necessary to assess whether the intervention was delivered as intended (fidelity). Fidelity has been defined as 'how far those responsible for delivering an intervention actually adhere to the intervention as it is outlined by its designers' [14]. The degree to which intervention fidelity needs to be monitored within RCTs will largely depend upon the extent to which the trial protocol prescribes standardisation.<br>Fidelity is also referred to as compliance or adherence. |
| Blencowe et al | 2016 | Surg | PP | n/a | NR | Monitoring how surgical interventions are actually delivered in a trial (fidelity) is essential to inform the interpretation of results and subsequent implementation of interventions in practice. Three possibilities for |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | recording and reporting fidelity were identified: the intervention, component or step is not delivered at all; an intervention, component or step from another trial group is delivered instead; or an entirely different intervention, component or step is delivered. |
| Boden, et al | 2016 | PT | MMS Nested in RCT | Nested MMS to investigate Rx differentiation, receipt, and enactment. Interviews, recordings, thematic analysis. | Borrelli NIH-BCC | the intervention is differentiable from standard care, is memorable, and adhered to by the patient. Treatment fidelity has four components [13]: (i) Integrity; was the treatment delivered as intended? (ii) Differentiation; did two treatments differ from one another as intended? (iii) Receipt; does the patient understand the treatments provided and are they equipped to perform them as intended?; and (iv) Enactment; does the patient enact the learnt skills and perform the intervention as intended? |
| Boden, et al | 2018 | PT | PTCL RCT | TIDIER | NR* | *Described elsewhere? Boden I, El-Ansary D, Zalucki N, Robertson IK, Browning L, Skinner EH, et al Physiotherapy education and training prior to upper abdominal surgery is memorable and has high treatment fidelity: a nested mixed-methods |
| Bowen, et al | 2018 | SLP | FS RCT | Intervention details were recorded in speech therapy notes and retrieved following completion of follow-up. | NR | NR |
| Boyd | 2017 | PT OT | PTCL RCT | Manual, checklist for fidelity, observations/video analysis | NIH-BCC | study design, training of intervention providers, treatment delivery, treatment receipt and enactment |
| Boyle, et al | 2009 | SLP | RCT | evaluated modes of service delivery | NR | NR<br>evaluated modes of service delivery with MRC guidance to ensure fidelity |
| Brecke nridge | 2015 | OT | PP | NA | Perplko va | Fidelity is achieved by ensuring: practitioner adherence to specified procedures; practitioner competence in delivering the intervention; and clear differentiation between interventions: that is, the intervention is substantively different to other approaches with similar goals. |
| Broekh uizen, et al | 2012 | LI PA | PE RCT | MI fidelity) was assessed by two MI experts, following the Motivational Interviewing Treatment Integrity code. Process evaluation. | RE-AIM | NR<br>including dose and fidelity monitored. Fidelity is defined as the extent to which the intervention was implemented as intended. To what extent was face-to-face counselling delivered as planned by MI guidelines?<br><br>DOSE NOT PART OF FIDELITY |
| Brogan | 2019 | SLP | PP | N/a | NIH-BCC | study design, training providers, delivery of treatment, receipt of treatment, and enactment of treatment skills |
| Brogan | 2018 | SLP | PP | N/a | NIH-BCC | study design, training providers, delivery of treatment, receipt of treatment, and enactment of treatment skills |
| Bronars , et al | 2017 | PA LI | FidSt | Various: Protocol deviations manually recorded using a checklist form. Interviews, audits, etc. | NIH-BCC | design focuses on the methodological processes that ensure the study adequately assesses the proposed hypotheses in relation to a theoretical framework.<br>Training was designed to ensure satisfactory delivery of the intervention to study participants. Trainings were tailored to account for different backgrounds and past training experiences of the FHPs.<br>Treatment fidelity pertaining to treatment delivery includes ensuring that the content and dose are consistent as well as adherence to the manual.<br>Receipt ensures that participants received and understood the treatment provided.<br>Enactment assessment and monitoring of participant behaviour outside of the intervention |

| | | | | | | |
|---|---|---|---|---|---|---|
| Bryant, et al | 2014 | PT | FidSt | This audio recording was reviewed by the site psychologists against specified criteria for fidelity to the content and quality of its delivery | NIH-BCC | treatment fidelity, a term that refers to the consistent and reliable delivery of interventions. achieving training fidelity: a specific intervention cannot be delivered until those delivering it have learned to do so in a standardized way. |
| Burkart | 2017 | PA | PE | semi-structured questionnaire completed by trained data collector observing sessions. Participant accelerometers. | Durlak and Dupre | Intervention fidelity (i.e., adherence, compliance, integrity). degree of which the intervention was delivered as originally planned |
| Busse, et al | 2017 | PT | RCT | Fidelity of the physical activity intervention was measured using a combination of self-report checklists, independent analysis of audio recordings, and a self-assessment completed by the intervention coaches. | Not reported, but see Quinn et al, 2016, for description. | NR |
| Butel, et al | 2015 | LI<br>PA | cRCT | fidelity monitoring, which documented how well intervention components were being implemented, assessed intervention adaptations (expected given vastly different intervention site contexts), and identified ways to improve intervention delivery. | NR<br><br>Cites Gearing | NR |
| Bysterveldt et al | 2010 | SLP | SS | Analysis of videotaped sessions. | NR | NR |
| Campbell, et al | 2015 | PA | FidSt | Process evaluation, quantitative and qualitative | NR | Fidelity is whether the intervention is delivered as expected whether and how variations in delivery occurred. To what extent was the intervention delivered as planned? To what extent was the intervention delivered as planned? IN what ways, if any, did the teachers amend the programme? and<br>What were the reasons for any amendments? |
| Carlson, et al Lifestyle intervention for adults with spinal | 2019 | RH<br>LI | RCT | Interveners received 30 hours of standardized training, were assessed monthly in session delivery using a standardized rating scale, and attended weekly troubleshooting | NIH-BCC | NR |

| | | | | | | |
|---|---|---|---|---|---|---|
| cord injury | | | | meetings to mitigate protocol drift | | |
| Carlson, et al | 2019 | PA SCI | RCT | multi-faceted treatment fidelity plan : standardized training and check ins. | Bellg NIH-BCC | To ensure that the intervention protocol was properly implemented, we adhered to a multi-faceted treatment fidelity plan consistent with guidelines for monitoring complex interventions. According to this plan, interveners received 30 hours of standardized training, were assessed monthly in session delivery using a standardized rating scale and attended weekly troubleshooting meetings to mitigate protocol drift.50–52 All interveners were blind to the study hypotheses and design. |
| Carroll | 2007 | Fidelity | PP | n/a | Carroll | 1.Adherence: whether "a program service or intervention is<br>being delivered as it was designed or written.<br>2.exposure or dose: the amount of an intervention received by participants (frequency and duration as specified in protocol…)<br>3.quality of delivery: the manner in which therapist delivers intervention. E.g.: adherence achieved but the intervention delivered badly. It concerns whether an intervention is delivered in a way appropriate to achieving what was intended. If the content of an intervention is delivered badly, then this may affect the degree to which full implementation is realised.<br>4.participant responsiveness: how far patients respond to or are engaged by an intervention.<br>5.programme differentiation: Identification of an intervention's essential<br>components.<br><br>CARROLL: **Adherence**: content, coverage, frequency, duration. Adherence is essentially the bottom-line measurement of implementation fidelity. If an implemented intervention adheres completely to the content, frequency, duration, and coverage prescribed by its designers, then fidelity can be said to be high. The **content** of the intervention may be seen as its 'active ingredients'; the drug, treatment, skills, or knowledge that the intervention seeks to deliver to its recipients. Subcategories of adherence concern the frequency, duration, or coverage of the intervention being delivered, i.e., what is more broadly defined as "**dose**" Identifying these essential components also provides scope for identifying adaptability to local conditions. Moderators include complexity, delivery qual… |
| Case-Smith, et al | 2011 | OT | PS | Customized Fidelity measure | NR | NR |
| Case-Smith, et al | 2011 | OT | UCT | | Adherence to Rx principles. | NR<br><br>Fidelity: Adherence to Rx principles. |
| Casey, et al | 2018 | PA PT | PTCL RCT | Treatment fidelity will be assessed by a health psychologist who is highly experienced in the delivery of group psychological interventions for chronic pain using ACT. All eight | Borrelli (NIH-BCC) | Fidelity: Fidelity to protocol. |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | ACT sessions from one treatment group will be audio recorded and reviewed. An ACT treatment fidelity tool that has been modified for chronic pain will be used to evaluate the intervention. Treatment fidelity of the physiotherapy components of the trial will be assessed review the checklists and notes at monthly intervals | | |
| Casida, et al | 2012 | Surg RH | PS | Log to check compliance to protocol (nurses) | NIH-BCC | Treatment fidelity consists of five components of design (for accurate testing of the study's clinical process), training (of those researchers providing the intervention), delivery (to monitor the intervention ac- curacy), receipt (to ensure that the patient is able to perform the intervention), and enactment (the ability to perform the intervention in real life settings) |
| Chesworth, et al | 2015 | RH | FidSt | Assessment of nurse's logs Focus on fidelity of treatment delivery on study. | NIH-BCC | These recommendations include five areas of treatment fidelity: study design, provider training, treatment delivery, treatment receipt, and enactment of treatment skills. Focus on fidelity of treatment delivery on study. |
| Chewning, et al | 2019 | PA | RCT | Master teacher's observation/a 40-item dichotomous checklist. | Borrelli, NIH-BCC Re-AIm | "Also, demographic information, fidelity of course delivery, course size and attendance, and fidelity of course receipt and enactment in terms of participant tai chi home practice was collected. Trained researchers collected physical, confidence and executive function data at community sites using specific stations for each measure with participants systematically rotating between these stations." |
| Clark, et al | 2014 | RH | PP RCT | 30 hours of standardized intervener training; (b) one monitoring session per month in which each intervener's adherence to the protocol was assessed using a specialized rating scale; and (c) weekly supervisory protocol adherence meetings. | NR | A balance between protocol adherence and clinical judgment must be maintained. We achieved this requirement through two strategies. First, the six intervention modules served as a toolbox demarcating the core components of the intervention and the range of appropriate content and provided troubleshooting guidelines. Second, interveners were expected to adhere to a set of theoretically grounded, overarching principles related to pressure ulcer risk when tailoring the sessions to be participant- and situation-specific |
| Cooke, et al | | RH | PTCL RCT | training and a standardised procedure manual (detailing protocol, plans for dealing with intervention fidelity issues, and monitoring the delivery and receipt of the intervention. | TDF<br><br>Cites Spillane | NR<br><br>Fidelity: Theoretical Domains Framework, delivery, and receipt of the intervention. |
| Cox, et al | 2018 | PT | FS PILOT RCT | subjective description of case notes by the study | | Fidelity is usually interpreted as the consistent delivery of intervention components.112 However, physiotherapy interventions do not always benefit from consistency as |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | team. The co-applicant physiotherapists rated each physiotherapy session for optimisation using data from the study documentation. Optimisation assessment of whether or not any adaptations had been made and whether or not the session had been completed | | they involve the revision of treatment plans to account for changing and uncertain experiences.113 Evidence suggests that the fidelity to the treatment theory is more important114–116 than consistent delivery of the intervention. We therefore evaluated treatment optimisation to the patients' needs and capabilities. Adherence to the prescription. Assessment of overall treatment optimization. It follows that optimisation, or 'fidelity of function' is more an appropriate construct for assessing implementation quality than 'fidelity of form',114 ensuring congruence with the intervention theory. |
| Cristiansen, et al | 2010 | PT PA | RCT | NR | NR Cites Pereplechikova | NR |
| Cutchin, et al | 2009 | OT | FS RCT | Manual. Feasibility is being measured by the rate of recruitment, representativeness of the study sample, and retention rates. | NIH-BCC<br><br>Cites Frank | fidelity in design, training, delivery, receipt, and enactment.<br><br>Frank https://doi.org/10.1177/0733464807308621 |
| Dean, et al | 2018 | PA RH | PILOT RCT | attendance registers, accelerometery, exercise 'homework' diaries, trainer completed session checklists and video analysis of (early, middle, and late programme) training sessions | NR | NR<br><br>No definition of fidelity: Adherence to protocol/manual. |
| Dean, et al | 2016 | RH stroke | PTCL Pilot RCT | attendance, accelerometery, exercise diaries, session checklists and video analysis of selected training sessions in each programme. | NR | NR<br><br>Fidelity: adherence to the intervention manual by participants and trainers |
| Deary, et al | 2018 | SLP | Pilot RCT | Sessions audio recorded and structured analysis of the content of intervention for treatment fidelity and inter-treatment contamination. monitoring the content of the sessions via real-time clinical supervision and retrospective content analysis of contemporaneous case notes. | NR | NR<br><br>Fidelity of delivery, protocol violations. |
| Deary,et al | 2018 | SLP | RCT DS | assessed by monitoring | NR | NR<br><br>No results of Fidelity Assessment. |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | the content of the sessions via real-time clinical supervision and retrospective content analysis of TM's contemporaneous case notes. | | |
| Desveaux, et al | 2016 | PT PA | PTCL RCT | Fidelity will be monitored throughout the study through semi-annual check-ins with community facilities. Facilities will explicitly outline the operationalization of the intervention at their respective facility, including frequency, duration, supervision, attendance monitoring, and individual program components. | NR | NR |
| DeVito, et al | 2011 | RH | FidSt | n/a | Custom | Delivery: the extent to which the intervention is delivered as intended; (b) receipt: the extent to which the intervention is received as intended; and (c) technology acceptance: the extent to which the participant has positive perceptions, attitude, and intention to use a system. Relationships purported in the framework include the following: (a) Intervention fidelity extends beyond delivery to include receipt and technology acceptance (perceived ease of use, perceived usefulness, attitudes toward use, and intention to use). (b) There is a reciprocal relationship between delivery and receipt (i.e., qualities of delivery affect receipt and vice versa). And (c) human factors (technology acceptance) moderate the relationship between delivery/receipt and ultimate adoption (use of technology).<br><br>Measurement of delivery typically includes assessing whether all the intervention components and activities were delivered and implemented in the proper manner. Adoption is the extent to which the individual participant uses the technology-based behavioural intervention. It is akin to terms such as intervention usage, utilization, and intervention dose and should not be confused with the use of the term for describing diffusion of innovations (how new ideas and technologies spread among groups. Enactment is the extent to which the participant performs the behaviours that the technology-based behavioural intervention is intended to promote (e.g., follow an exercise regimen, monitor health indicators). Because adoption moderates the relationship between the intervention and treatment effects (enactment), it is important to quantify both to determine the strength by which one can conclude that the intended outcomes were indeed due to the use of intervention. Because neither adoption nor enactment measures how well the intervention was delivered as conceived and planned, they are not included as components of intervention fidelity. |
| Di Rezze | 2012 | RH PT | NR | | IFF | evaluating the adherence of the therapist to an intervention and differentiating therapist behaviours |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | OT | | | | between interventions. Therapist behaviours "client responsiveness," a concept that is also applicable to paediatric rehabilitation. Such items examined client or client-therapist interaction behaviours. Examples include, "how receptive and/or engaged was the client during the session? Fidelity includes therapy process (e.g., assessment) and delivery process (e.g., rapport). |
| Di Rezze et al | 2014 | RH | NR | n/a | IFF | actors that mediate intervention fidelity focus on therapist adherence to delivering the intervention as expected and include only therapist behaviours. Moderating factors of intervention fidelity examine concepts beyond the therapist's adherence to the intervention. Examples include therapist quality of intervention delivery, client attributes independent of the therapist, and client attributes dependent on the therapist (i.e., therapist-client interaction). |
| Di Rezze, et al | 2013 | PT OT | FidSt PP | Paediatric Rehabilitation Observational measure of Fidelity (PROF) developed for study. Scoring video of Rx. | IFF | The fidelity measure created to support this process was based on a conceptual model, the implementation fidelity framework (IFF). In this model, therapist adherence in delivering the intervention mediates intervention fidelity. potential moderators may occur before intervention delivery, such as the complexity of an intervention (i.e., comprehensiveness of policy) and defining strategies to facilitate therapy better. Quality of therapy delivery and participant responsiveness moderators were related to examining the deliverer (i.e., therapist) and recipient (i.e. client) within an intervention session. Adherence: details of content, coverage, frequency, and duration. |
| Di Rezze,et al | 2014 | RH | PP | CS | NR | NR |
| Dillon, et al | 2018 | | PTCl PE | quantitative (fidelity checklist score, number of completed sessions, survey data and a habit formation scale), as well as qualitative (open responses from program staff and semi-structured interviews with study participants) data | MRC | program adherence (fidelity): How well did the specialists deliver the program to participants?<br>complete delivery (dose delivered) To what extent were all of the intended components of the v-LiFE program delivered to participants?<br>participant receipt (dose received: To what extent were participants engaged/satisfied with the v-LiFE program?<br>participant enactment: To what extent were the participants completing the prescribed activities? |
| Drew, et al | 2016 | CI | CS PP | Evaluation of recordings. | NR | fidelity of the delivery of the intervention in line with the theoretical framework on which it was developed |
| Dunn, et al | 2017 | OT LI/HP | FidSt | Hand coding of observations/ transcripts of randomly selected sessions against checklist made for study, based on Gearing and components of OPC. *Blind coding/assessment of fidelity in Rx sessions.* | Geariin g and Bellg (NIH-BCC) | Gearing Model of NIH-BCC. These authors described four key elements of fidelity: (a) intervention design, (b) intervention training, (c) monitoring intervention delivery, and (d) monitoring receipt of intervention. Researchers must use a clearly outlined theoretical framework, program goals, and a consistent methodology to measure the fidelity of an intervention design.<br>ESSENTIAL actions of the intervention and UNIQUE actions of the intervention.<br>It is also important to monitor specific aspects of intervention receipt, including dose and participant comprehension, resistance/acceptability, attendance, and/or knowledge before and after the intervention.<br>assessed intervention receipt by monitoring participant attendance and participation throughout the study. We queried parents regarding acceptability of OPC for their families and their goals via post intervention questionnaires. |

| | | | | | | |
|---|---|---|---|---|---|---|
| Elinder, et al | 2012 | PA | UCT | Process evaluation. Fidelity to school action plans was evaluated through interviews with health teams guided by a checklist. | NR<br><br>Cites Fraser | Fidelity to the programme as a whole was assessed in relation to whether schools had implemented all components in the programme according to the logic model.<br>Fidelity, defined as the extent to which a programme adheres to its programme theory |
| Eyre, et al | 2016 | CR | PTCL Pilot RCT | A customized, 13-item intervention fidelity checklist. audio recordings by intervention facilitators and then reviewed and coded (using the fidelity checklist) by two *independent researchers.* | NR | NR<br><br>Fidelity: Adherence by facilitators to the intervention protocols |
| Eysenbach, et al | 2019 | PA RH | PTCL RCT | "expert auditing of calls, monitored study website, coaching call notes. Monitoring all phases. | NIH-BCC | "study design, provider training, delivery of treatment, receipt of treatment, and enactment of treatment. " "**study design** area focuses on practices that ensure study procedures and implementation are in line with current theory and clinical processes. Study design fidelity goals include ensuring that conditions are congruent with relevant theory and practice, ensuring equivalent treatment dose within and across conditions, and planning for implementation setbacks." **Provider training: "**strategies that address preparation for uniform delivery of treatment by providers/coaches." "standardized training; this ensures that training is conducted similarly for all providers. " "Another goal of provider training is minimizing drift in provider skills. **"Delivery**: "Fidelity of treatment delivery focuses on ensuring the intervention is delivered as intended. Many of the concerns within delivery of treatment overlap with strategies for training and study design, including controlling for provider differences and adhering to created protocols; however, this area further addresses differences within treatment conditions and minimizes contamination " "**treatment receipt** involves strategies and monitoring of a participant's ability to understand and adopt treatment-related behavioural skills and cognitive strategies. " "participants' comprehension and ability to utilize digital media in delivering content and tracking goals" "**enactment** of treatment described as strategies aimed at monitoring and improving participant ability to perform treatment-related behavioural skills and cognitive strategies in relevant real-world settings." |
| Ferrante, et al | 2019 | PT | PTCL RCT | systematic process, specific therapist training for motor learning principles and techniques as well as administration of outcome measures, establishment of interrater reliability with outcome measures, treatment session observation, critique of participant | Gearing | Key components of intervention fidelity include a well-designed framework, therapist training, monitoring of treatment delivery, and monitoring of treatment receipt (by participants) |
| Flynn<br><br>SMART | 2018 | Pain RH | RCT | Role-playing ensuring that team members understand the | NIH-BCC | NR |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | protocol. All contact with participants is scripted and randomly and regularly reviewed by the principal investigator. Treatment delivery is monitored by the interdisciplinary pain management centre clinical team leader | | |
| Foster, et al | 2016 | RH PT | FS PILOT RCT | PTs recorded full details of the advice and treatments, number and mode of treatment sessions, any non-attendance, acupuncture points used, any sensations during acupuncture treatments and any adverse events on specifically designed CRFs. Audited against the PT clinical notes to ensure accuracy and to determine protocol adherence. | NR | NR<br><br>Fidelity to protocol<br><br>Audited against the physiotherapists' clinical notes to ensure accuracy and to determine protocol adherence by participating physiotherapists before collation by the research team in order to fully describe the interventions delivered. Where protocol deviations were noted, these were discussed with the physiotherapists involved in order, whenever possible, to enhance adherence to the agreed intervention protocol. |
| French, et al | 2015 | PR | SR | five-area (eight-element) intervention fidelity tool, | NIH-BCC | Intervention fidelity (study design, training of providers, delivery of treatment, receipt, and enactment of treatment?<br>Study design: was the guideline or protocol used to guide the study published and was it clearly identified? were standardized or validated tools used to measure patient reported outcomes?<br>Training of providers: formal training of providers related to the guideline or protocol used, and was an intervention manual used to guide providers?<br>Delivery of Treatment: core treatment interventions consistent with the guideline or protocol used to develop the intervention manual and/or to guide the study? "was there assessment of response to treatment at specified timeframes?<br>Receipt of Treatment: Was there any mention and/or measurement of concordance of patient and provider understanding of the problem and/or treatment recommendations?<br>Enactment: patient's ability to engage in the treatment recommendations in daily life |
| French, et al | 2011 | PA | PTCL | Manualized intervention. Taped intervention sessions are being analysed to quantify the extent to which each intervention technique was delivered as specified by the intervention manual. Selection of patients were interviewed immediately after receiving the | Bellg (NIH-BCC) | Provider fidelity of delivery to the intervention manual and maximise recipient fidelity of receipt and enactment of the intervention techniques. If interventions are not delivered or received as intended (i.e., as per protocol) then it is difficult, if not impossible to be certain that the results can be attributed to the intervention itself. fidelity in relation to delivery and receipt of the intervention |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | intervention to monitor intervention receipt. | | |
| Fuentes, et a; | 2014 | PT | RM | Therapist adherence was based on how closely the therapists followed the experimental protocol assessed by videotaping all treatment sessions, of which 28 (20%) were randomly selected for evaluation . Two research assistants not involved with the study separately rated each session regarding treatment fidelity. | NR | NR |
| Fuentes, et al | 2014 | PT | expContSt | Therapist adherence was assessed by videotaping all treatment sessions, of which 20% were randomly selected for evaluation. Two research assistants not involved with the study separately rated each session regarding treatment fidelity. *Independent assessments.* | NR<br><br>Cites Moncher | Therapist adherence was based on how closely the therapists followed the experimental protocol. |
| Furnes et al | 2018 | PA | PE | Process evaluation of fidelity | MRC | NR |
| Galaviz, et al | 2014 | PA | SR | RE-AIM criteria in checklist | RE-AIM | the duration and frequency of the intervention, the extent to which the protocol was delivered as intended, and the cost of delivery. Additional implementation indicators used include whether articles reported the theoretical framework of the intervention, the consistency of implementation across settings and delivery agents, the degree to which the participants received intervention components, and the use of qualitative methods for measuring implementation. |
| Gibson, et al | 2016 | PR | SR | five-area (eight-element) intervention fidelity tool, | NIH-BCC | Study Design, Provider Training, Treatment Delivery, Treatment Receipt, and Treatment Enactment |
| Giesbrecht, et al | 2017 | OT<br>PT<br>RH | RCT<br>FS | Trainers indicated any protocol deviations on the checklists and completed a trainer post-treatment evaluation form after EPIC | NR | Adherence to protocol |
| Gladwell, et al | 2016 | PT | QualStdy | Semi-structured interviews | Carrol | adherence and moderating factors. Adherence is defined as “how far those responsible for delivering an intervention actually adhere to the intervention as it is outlined by its designers”11 and includes the |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | subcategories of treatment content, coverage, frequency, and duration. Moderators are factors that influence the degree of fidelity with which an intervention is implemented and include intervention complexity, facilitation strategies, quality of delivery, and participant responsiveness to a treatment program. |
| Glombiewski, et al | 2010 | CP<br>BC | RCT | Manual, videotape analysis. | NR | All therapists were trained in a standardised treatment plus biofeedback and were supervised weekly by licensed cognitive-behavioural therapists and supervisors. Treatments were described in detailed session-by-session protocols to maintain treatment fidelity. Selected videotapes of therapy sessions were discussed during supervision contacts to ensure adherence to protocols. |
| Glombiewski, et al | 2010 | Pain<br>RH | | Selected videotapes of therapy sessions were discussed during supervision contacts to ensure adherence to protocols. | Fidelity to manuals | No theory reported.<br>"Treatments were described in detailed session-by-session protocols to maintain treatment fidelity." |
| Godecke, et al | 2015 | SLP | FidSt | Therapy fidelity was measured according to the TIDiER statement with treatment adherence and treatment differentiation monitored throughout the trial. sessions were reviewed by the independent therapy fidelity monitor and feedback was provided to therapists as part of the ongoing adherence to the treatment protocol. | NR | NR |
| Godfrey, et al | 2016 | PT | PTCL | sessions from every PT will be rated in terms of adherence to the manual and checklist/modified fidelity measure | NR | Fidelity to manual/protocol. |
| Golos, et al | 2011 | OT | qRCT | consultation and monitoring sessions were documented using forms for the purpose of determining treatment fidelity. Customised Fidelity measure | NR | NR<br>No results of Fidelity assessment. |
| Graham, et al | 2018 | PT<br>PA | PTCL<br>cRCT<br>FS | Observations of the delivery of a sample of the training will be conducted in each intervention home. Semi-structured interviews with staff and trainers. Process evaluation. a trial-specific observational tool designed to | NR | Intervention fidelity and adherence (training, content, delivery, attendance |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | record instances of care behaviours reflective of skills (that could have been) learned during SCTP training will be developed and tested in more than one site. | | |
| Granbom et al | 2019 | OT | PTCL | demonstrate the last session's exercises to the OT on the following session and show the logs of activities. Sessions are audio-recorded, and study team members review 10% of the sessions. bi-weekly meetings with the OT and PI case presentations evaluated in terms of fidelity to intervention delivery | NIH-BCC | "The fidelity plan, based on the National Institutes of Health (NIH) Behaviour Change Consortium, addresses fidelity through design (distinct program based on theory), training (using established LiFE program training, home safety training, and program manual), delivery (reminder calls the night before sessions), engagement (records of home sessions by date and duration), and receipt (completing checklists on program engagement)." |
| Granbom, et al | 2019 | OT | PTCL<br>PS<br>RCT | Sessions are audio-recorded,<br>and study team members review 10% of the sessions.<br>bi-weekly meetings with the OT and (PI) case presentations evaluated for fidelity to intervention delivery. | NIH-BCC | addresses fidelity through design (distinct program based on theory), training<br>(using established LiFE program training, home safety training, and program manual), delivery (reminder calls the<br>night before sessions), engagement (records of home sessions<br>by date and duration), and receipt (completing checklists on program engagement). To assure enactment,<br>participants in the LIVE LiFE group demonstrate the last session's exercises to the OT on the following session and show the logs of activities. |
| Guaglianno, et al | 2019 | PA | FS<br>RCT | Participant questionnaires. | NR | NR |
| Gunn, et al | 2018 | | | fidelity of a random sample of a minimum of 25% of the delivered sessions will be assessed using audio recordings of the session. This sample will include at least two recordings of each session type. Checklist used, two team members not Rx. | NIH-BCC<br><br>MRC | NR<br><br>Process evaluation for Fidelity. |
| Hahne | 2011 | PT | PTCL<br>RCT | comprehensive treatment manual, initial training of PTs, a monthly teleconference involving all treating PTs, clinical notes directing PT along decision making algorithms, reviewing the clinical notes of | NR<br><br>Cites Borrelli | NR |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | every participant at three points during their treatment program, and the use of standardised participant information sheets are methods chosen to ensure that all participants receive treatment from PTs that is standardised, accountable, and reproducible. | | |
| Hand, et al | 2018 | PT OT | SR | n/a | Carroll<br><br>Dane and Schneider. | Fidelity is the faithfulness of an intervention to its underlying therapeutic principles and clinical guidelines. Fidelity consists of five key components: (1) adherence, (2) quality of delivery, (3) exposure, (4) participant responsiveness, and (5) program differentiation. Adherence refers to the extent to which program components are delivered as intended, whereas quality of delivery is a subjective aspect of treatment delivery that extends beyond delivery of prescribed content. For example, aspects of quality of delivery may include clinician enthusiasm or attitude. Exposure refers to the number, length, or frequency of intervention sessions or the frequency with which intervention techniques are implemented. Participant responsiveness includes participants' judgments about the outcomes and relevance of an intervention (Carroll et al, 2007) and is a key aspect of intervention fidelity. Finally, program differentiation refers to how the intervention being delivered is different and distinguishable from other interventions<br>. |
| Hankonen, et al | 2016 | PA | PTCL cRCT | Intervention facilitators keep track of components delivered, as well as the quality of delivery (e.g., interaction elements), by filling in a self-assessment form after each session, to assess whether the intervention was delivered as intended and to ensure high fidelity | NIH-BCC | receipt and use of intervention materials (e.g., use of workout and activity break videos) and enactment of the BCTs taught in intervention classes |
| Harris, et al (Nauta) | 2013 | PA | PS, FS | Inferred from proxy. | NR | "Another strength of the study was that both interventions were implemented by highly skilled interventionists with years of training in their fields suggesting high levels of treatment fidelity by those implementing the intervention and assessment." |
| Hart, et al | 2012 | RH | PP | n/a | NR | Fidelity may be defined as the extent to which the core components of treatment have been delivered as intended. concepts of treatment receipt and treatment enactment. Treatment receipt refers to the extent to which the patient understands the strategies or techniques taught and demonstrates the capacity to use them. For this purpose, one could administer pre- and posttreatment tests of knowledge related to treatment. Treatment enactment, which has to do with whether the participant actually uses the learned strategies in day-to-day life, is more challenging to measure but could be |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | ascertained using self-report and proxy report instruments given at some point after the trial. |
| Harwood, et al | 2018 | PT OT HP/LI | PTCL FS | Records of therapy sessions undertaken, and self-directed activity will be examined, and video-recorded therapy sessions will be assessed qualitatively for fidelity. Mechanisms of impact and contextual factors include engagement and adherence. | NR | We will undertake a process evaluation, studying fidelity, understanding mechanisms and context. Implementation (delivery of intervention), includes fidelity (quality of delivery) and dose (quantity of delivery). |
| Healey, et al | 2018 | PA | FS RCT | Audio recording of intervention sessions and intervention fidelity checklist specifically developed for the trial to assess whether components of the consultation intended to be included | Gearing | fidelity – the degree to which the intervention is delivered as intended. which elements of this intervention were delivered. |
| ***Hildebrand*** | | | | | | |
| Hill, et al | 2014 | RH | cRCT | Therapists complete the online training programme. Patient's unit lists are checked. education and training are repeated for new staff on the unit. | NR | NR<br><br>Adherence to protocol |
| Hinckley, et al | 2013 | SLP | NR | n/a | Moncher and Prinz | Treatment integrity refers to how well a treatment condition was implemented as planned. treatment differentiation refers to whether the treatment conditions being studied differed from each other sufficiently so that the intended manipulation of the independent variable can be assumed to have occurred. Both of these concepts are important to consider because it is possible to administer a treatment as planned without differentiation from a comparison treatment or to successfully differentiate two treatments in a research study without implementing the treatment with a high degree of integrity |
| Hoekstra,et al | 2018 | PA (MI) | | n/a | MICAS | It is expected that the MICAS contains four subscales related to the factors: Acceptance (10 items), Partnership (8 items), Evocation (3 items), and MI non- adherence (5 items). |
| Hofman, et al | 2013 | PA PT | PTCL RCT | BET is delivered by the same BET-therapist, who will be intensively trained in the use of the program and will receive a detailed trainer manual. Regularly, announced visitations during the Rx | NR | high treatment integrity. The latter refers particularly to a potential mixture between the study group and the control group through a communication about intervention contents between patients and therapists within each rehabilitation centre, as well as individual deviations from the treatment protocol. |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | phase will occur in order to assure adherence to the treatment protocol. Individual deviations from the Rx protocol will be recorded. | | |
| Holland, et al | 2018 | SLP | FS | Video and audio footage assessed for fidelity with MIS Checklist. | | intervention maintains fidelity to the core principles of the therapy. session number, duration, and content; therapist background, training, and support. |
| Hollands, et al | 2013 | PT | | Adherence to the intervention by therapists will be assessed during their involvement in the trial by A. Wimperis and K. Hollands through video observation at weeks 2 and 6 of each therapists' first treatment period. Further training for the therapist will be provided, if necessary, to improve compliance with treatment protocols. | NR | NR |
| Holt, et al | 2018 | PA | RCT | *assessed by facilitator talk time, checklist and direct observation of the facilitator behaviour and conduct at sessions. | NIHBCC Linnan and Steckler's process evaluation framework | **Design**: Ensure the same treatment dose within conditions. Ensure an equivalent dose across conditions. Plan for implementation setbacks.<br>**Training:** Standardise training. Ensure provider skill acquisition. Minimise 'drift' in provider skills. Accommodate provider differences<br>**Delivery:** Control for provider differences. Reduce differences within treatment.<br>Ensure adherence to treatment protocol. Minimise contamination between conditions.<br>**Receipt:** Ensure participant comprehension. Ensure participant ability to use cognitive skills. Ensure participant ability to perform behavioural skills.<br>**Enactment:** Ensure participant use of cognitive skills. Ensure participant use of behavioural skills. |
| Hosseini, et al | 2018 | PA Tai Chi | RCT | We assessed the treatment fidelity of study by a standard guideline (Bellg et al, 2004). To ensure the exercise program was administered by the same frequency and duration of sessions for all participants. | NIH-BCC | NR |
| Hurd, et al | 2017 | PT | PTCL RCT | weekly teleconference meetings, in which the documentation from training is reviewed, and discrepancies addressed. Video | NR | NR |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | recordings of a full training session are compared periodically. Finally, each therapist visits the other centre at least once a year to observe and discuss the training. | | |
| Hurley, et al | 2019 | PT | cRCT | PT behaviour was assessed during delivery of the SOLAS intervention by audio coded by one blinded expert rater and physios' self-reported checklists to evaluate fidelity to the intervention content | Borrelli | NR |
| Hurley, et al | 2016 | PT | PTCL cRCT | direct observation and audio recording by researcher, PT self-report to assess the content and quality of treatment fidelity during the trial. interviews with intervention PTs .Fidelity will be assessed and reported by separate evaluators from the outcome evaluators. | | |
| Jafar, et al | 2016 | LI | FS | Process evaluation | NR | fidelity defined as the proportion of a) the planned orientation sessions delivered to physicians and nurses, b) the prescription of FDC to eligible participants, c) the delivery of MC to eligible participants, and d) the telephone follow-ups. |
| Jago, et al. | 2015 | PA | PE RCT FS | Process evaluation interviews | NR | dose delivered, the reach (or number of people who receive the intervention), fidelity (extent to which the intervention was delivered as planned), implementation (how well the programme was implemented) and context which provides critical information on the environment in which the programme was delivered |
| Jaka, et al | 2016 | LI | SR | Modified NIH-BCC Fidelity checklist. | NIH-BCC | treatment design, provider training, and treatment delivery, receipt, and enactment. |
| Johnston, et al | 2009 | OT | PP | Fidelity to manual | NR | NR<br><br>No definition of Fidelity |
| Johnston, et al | 2019 | SLP | RCT | Sessions were all video-recorded. Each session had a Treatment Fidelity Checklist that contained the essential elements of each session for the therapist. Therapists rated each element of therapist fidelity. A random sample of | NR | therapist integrity in the delivery of treatment as intended and (b) parent adherence to the delivered treatment |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | 10% of each therapist's video-recorded sessions was examined quarterly by an independent observer to determine inter-rater reliability. | | |
| Jørgensen, et al | 2012 | PT | RCT | Log-books on adherence, fidelity and context were held during the intervention. | NR | adherence, context and fidelity and the interplay between adherence and contextual events. implementation is reflected in the intervention dose received by the participants (i.e., adherence) and the fidelity (i.e. the quality of intervention delivery). |
| Karas | | MT PT | | | | |
| Katz et al | 2015 | Surg | PP | n/a | Cites Nelson and Cordray | Intervention fidelity refers to the extent that the intervention is delivered in an identical fashion to each subject.<br>To address these issues of intervention fidelity and learning curve, the surgical investigators should meet prior to the study launch in order to develop consensus on the precise surgical protocol, including the decisions to be made in the face of particular intraoperative findings. |
| Kearney, et al | 2006 | LI/HP | FidSt Pilot | Fidelity to protocol not assessed or recorded during trial. | Bellg<br><br>Leventhal and Friedman | Full execution of an intervention has been termed treatment fidelity, which has two components: integrity (the research team delivers the intervention as intended) and differentiation (the groups' exposure to the intervention differed as intended). Integrity includes treatment receipt (participants absorbed the knowledge and skills imparted in the intervention) and treatment enactment (participants used the skills in the manner intended. treatments be supported by a theoretical explanation for their mechanism of action so that they can be adapted without losing integrity, and failures can be explored in light of the theory. Differentiation of groups is the second major component of fidelity. |
| Keogh, et al | 2018 | PT | FS | process evaluations to<br>assess the fidelity to, and implementation of, intervention components | NR<br><br>MRC and Borrelli cited. | Fidelity guidelines suggest that following training, providers should be competent to deliver the intervention. Fidelity evaluations involve the assessment of what has taken place |
| Kerr et al | 2018 | PA | cRCT | Evaluation of participation (dose?) and enactment. | NR | NR |
| Kipplin g, et al | 2016 | PA | FS cRCT` | Process evaluation | NR | NR<br><br>reach and dose, some enactment |
| Lambert, et al | 2018 | PA | Pilot RCT | Delivery fidelity was tracked using Web usage statistics. Fidelity of receipt 5-point Likert response scale. assessed participants' perceived ability to use the intended BCTs by asking participants to rate their confidence in using specific BCTs. to assess enactment, asked participants if they had used | NR<br><br>Cites Borrelli (NIH-BCC) | Intervention fidelity was conceptualized and measured in the domains of design fidelity, training fidelity, quality/completeness of delivery, participant receipt, and enactment. fidelity of intervention delivery, receipt, and enactment (use of techniques). |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | specific BCTs related to BA in the last 2 months using a binary scale (yes/no) | | |
| Lambert, et al | 2017 | PA | SR<br>Fid | NIH-BCC criteria | NIH-BCC<br>MRC | Study Design, Provider Training, Intervention Delivery, Intervention Receipt and Enactment. Study Design is concerned with whether a study adequately tests its hypotheses in relation to its underlying theoretical and clinical processes. Provider Training involves standardizing training between providers and ensuring they are trained to clear criteria and monitored over time. Intervention Delivery involves assessing and monitoring differentiation (differences between the intervention and any comparison treatments), competency (skills set of provider), and adherence (delivery of intended components). Intervention Receipt refers to whether the intervention was understood and 'received' by participants and enactment refers to whether participants used intervention related skills in day to day settings |
| Lamdesman-Ramey, et al | 2019 | PT<br>OT<br>CIMT | PTCL<br>RCT | Videotaped sessions each week for every child and then scoring randomly selected 15 min segments | Standardised Fidelity of Implementation Measure (FIRM) developed for the trial. | Not reported in the protocol. Emailed lead author 06/10/19. |
| Lawford, et al | 2019 | PT | Case Study | Training facilitator audited audio-recordings of all consultations, and therapists self-audited 50% of consultations using a tool. | NR | NR |
| Lawrie, et al | 2018 | PA | Pilot RCT | a 2-week formal training in the execution and governance of the protocols, procedures for data collection and recruitment was given to the ward staff and the research assessors. See last column… | NR | NR<br><br>Fidelity defined by: "adherence to protocol (intervention fidelity).<br>to monitor and provide support to increase fidelity with the research protocol, a further visit after 3 months was undertaken and then regular video-conferencing calls were made every 2 weeks, and data were sent via protected email every 2 weeks to Oxford for storage and analysis" |
| Leeuw, et al | 2009 | RH<br>LI | PP<br>FidSt | Method of Assessing Treatment Delivery (MATD) aimed to determine protocol adherence and treatment contamination | MATD<br><br>Cites Pereplechikova | The assessment of treatment delivery consists of verifying the occurrence of essential components (protocol adherence) and the non-occurrence of prohibited protocol deviations (absence of treatment contamination) as well as verifying sufficient treatment differentiation<br>protocol adherence, referring to the degree to which specific treatment procedures are used by the therapists during actual delivery of treatment [8], competence, which is the skilfulness of the therapists delivering the treatment [8], and differentiation, signifying whether the |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | therapies differ from other treatments on several critical dimensions |
| Lenker, et al | 2010 | RH<br>OT<br>PT | PP | n/a | Moncher | Treatment fidelity is comprised of two components. The first is treatment integrity, the extent to which interventions are implemented as intended for the duration of a study (i.e., each participant receives the intended treatment). The second is treatment differentiation, the extent to which the differences between intervention and comparison conditions are maintained over the duration of study |
| Levy, et al | 2018 | PR<br>PA | PTCL<br>RCT<br>FS | NR | NR | Fidelity assessment: compliance with training protocol, monitoring of intervention delivery competency. |
| Li, et al | 2016 | RH<br>LI | FS<br>RCT | Fidelity check against competency/Rx checklist. | NR | NR<br><br>No definition of fidelity |
| Liu, et al | 2017, 2019 | PT<br>RH | PE<br>RCT | coordinators collectively trained at study initiation and annual collaborator meetings. on-site training as required. Day-to-day support by a clinical coordination team, neurologist and PT. A log of trial interventions kept by the coordinator for each participant for hospital and home visit activities. Intervention patients (with their caregivers) were encouraged to keep a daily log of rehabilitation activities for 30 days after discharge. | NR | NR<br><br>Fidelity: components implemented as per protocol |
| Liu, et al TaiChi | 2018 | RH<br>LI | PTCL<br>RCT | will be monitored through attendance records for each intervention component (peer support groups, health education seminars, and Tai Chi Ruler exercise) and personal records of the home practice of Tai Chi Ruler exercise kept by each participant. | NR | NR |
| Logan, et al | 2018 | RH<br>PT | PTCL | Evaluated using a trial specific SPIRES checklist that outlined all components of the functional standing frame programme intervention. PT will record the content of | MRC | The Medical Research Council guidance [43] recommends process evaluation and highlights the importance of capturing fidelity (whether the intervention was delivered as intended); dose (the quantity of intervention implemented) and reach (whether the intended target population comes into contact with the intervention, and how). |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | their sessions and adverse events in the Case Report Forms; an independent assessor will observe one intervention/one control group session at each of the four Stroke Rehabilitation Units at random timepoints during recruitment and complete a fidelity checklist (during qualitative interviews with PT. | | |
| Lotzke, et al | 2019 | PT | FS | Protocol developed and tested for feasibility Observation of treatment delivery by PIs. | MRC NIH-BCC | Fidelity is a process applied in the study design, training the provider, delivering of treatment, receipt of treatment and enactment of treatment skills Rx Receipt: assessing and optimising the degree to which the participant understands and demonstrates the knowledge provided by the intervention). |
| Lousada, et al | 2013 | SLP | RCT | Author and blind observation, checklist. | NR | "observational rating scale recording key elements: duration of session; target sound(s); type of reinforcement used; type of intervention; and main activities used." |
| Macleod, et al | 2018 | LI | FS | Programme implementation (by LCs) was estimated from a structured pro forma completed after every patient contact which recorded actual values or scaled ratings | NR | protocol adherence and acceptability |
| Malden, et al | 2018 | LI | PTCL | Intervention fidelity will be assessed using questionnaires and interviews with parents and practitioners, observation, and session delivery records. Intervention fidelity. measured by assessing the practitioner's logbook of the number of sessions conducted per week, and structure followed in relation to classroom guides. Observation of delivery will also be undertaken, as will qualitative interviews with practitioners and parents on completion of the intervention. | NR | NR |

| | | | | | | |
|---|---|---|---|---|---|---|
| Mars, et al | 2013 | CI Pain | PP FidSt | courses were audio recorded with the consent of participants and these recordings were used to assess and evaluate intervention integrity relating to the key elements prescribed in the COPERS facilitator's manual. | NIH-BCC | intervention delivery or integrity, defined as the monitoring and assessment of behaviours at the point of intervention delivery. The effectiveness of complex interventions may be dependent on the 'skills' of those delivering them. 'Skills' can be characterised by separate but related constructs of adherence and competence. Adherence is defined as: the extent to which a person delivers the essential content, delivery strategies and theories prescribed by the intervention designers and avoids activities proscribed by them. Competence refers to the level of 'skill' demonstrated by those delivering an intervention and may include the ability to respond appropriately to a wide variety of contextual cues. (DOSE NOT PART HERE). |
| Masterson-Algar | 2014 | OT | PP | n/a | Realist Evaluation, consolidated framework for implementation research (CFIR) | The programme theory incorporates four potential mechanisms through which fidelity within the trial can be investigated. These four programme theory areas are (1) the balancing of research and professional requirements that therapists performed in a number of areas while delivering the study interventions; (2) the OTs rapport building with care home staff; (3) the work focused on re-engineering the personal environments of care home patients; and (4) the learning about the intervention within the context of the trial and its impacts over time. |
| Masterson-Algar et al | 2017 | SLP | PTCL, PE RCT | Process evaluation: questionnaire, observations, interviews. | NR | process evaluation will be focused on investigating the quality of implementation of the PD COMMinterventions as well as adherence to the outcome evaluation protocol |
| Masterson-Algar, et al | 2018 | RH | PP CP | n/a | | researchers should investigate barriers and enablers to implementation by reviewing strategies in place to improve or support the fidelity of the rehabilitation intervention. The process evaluation should review strategies in place to measure 'dose delivered' and 'dose received'. Finally, participant's experiences and acceptability of the intervention should be investigated. To date, it is rare for research studies to provide intervention providers with clear guidance on how to assess which is the 'right amount' of tailoring |
| Masterson-Algar, et al | 2014 | OT RH | PP | N/a | CFIR | Consolidated framework for implementation fidelity (or research, CFIF or CFIR): adherence(intervention content, coverage, frequency, and duration), and 'moderating factors' that can potentially affect the degree of fidelity. These include the complexity of the intervention(s) to be implemented, facilitation strategies, quality of delivery and participant responsiveness: context, described as the culture of organizations, social behaviour/interactions among members and social structures, as an additional moderating factor |
| Mathews, et al | 2014 | PA | SR | REAIM | RE-AIM | Implementation refers to whether an intervention was delivered as intended in relation to protocol fidelity, provision of training and support for individuals delivering interventions may improve protocol fidelity. |
| Maxwell, et al | 2017 | RH PT | PTCL | Realist Evaluation to track how the implementation is working (including fidelity to the PFMT protocol). | NR | fidelity or variation to PFMT protocol (e.g., number and type of sessions) and the impact of any variations |
| Mayer-Davis | 2018 | LI | RCT | Coaches received training in motivational | NR | NR<br><br>fidelity of intervention delivery, |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | interviewing and problem-solving skills training and in the specifics of intervention delivery. Review of 10% of random selection of the audiotaped sessions for assessment of adherence to motivational interviewing principles using the motivation- al interviewing treatment integrity system. | | To ensure intervention fidelity, meaning that the intervention would be delivered as designed |
| McCarthy, et al | 2015 | PA | PE | Audiotaped sessions reviewed by independent assessor for adherence to MI principles. | NR<br><br>Cites Miller and Rollnick | The implementation of the intervention according to protocol, which includes fidelity, dose, and context. Evaluation of the fidelity to the intervention (how closely it was implemented as designed) focused primarily on the incorporation of MI principles into the counselling sessions. Essential elements covered, Adherence to principals<br><br>DOSE NOT INCLUDED IN FIDELITY: Evaluation of the dose of the intervention that was delivered consisted of examining the quantity or amount of intervention delivered to participants. |
| McDowell, et al | 2017 | CR PT | RCT | Various, including weekly phone calls with the research team to discuss individual patient treatment plans and regular training updates | NIH-BCC Borrelli | Design: "treatment fidelity practices relating to design ensure that a study adequately tests its hypothesis in relation to its underlying theoretical and clinical processes"<br>Training providers: "treatment fidelity of provider training involves standardising training between providers, ensuring that providers are trained to criterion, and monitoring and maintaining provider skills over time."<br>Delivery of treatment: "the assessment and monitoring of treatment fidelity during treatment delivery involves treatment differentiation (did the providers only deliver the target treatment and not other treatments), treatment competency (did providers maintain the skill set learned in training), and treatment adherence (delivery of the treatment components as intended)."<br>Receipt of treatment: "whether the treatment that was delivered to the participant was actually "received" by the participant."<br>Enactment of treatment skills "assessment, monitoring, and improving the ability of participants to perform treatment related behavioural skills and cognitive strategies in relevant real life settings |
| McPherson et al | 2018 | RH | | After each coaching session, the coach completed the fidelity measure, which was reviewed by a member of the research team with coaching expertise-Solution-Focused Fidelity Instrument (SFFI) | NR<br><br>Uses SFFI-Dumas et al, 2001 | NR<br><br>SFFI based on: Fidelity refers to delivering the intervention in a "comparable manner to all participants and is true to the theory and goals underlying the research" |

| | | | | | | |
|---|---|---|---|---|---|---|
| McPherson et al | 2019 | PA | PTCL RCT | Solution Focused Coaching Fidelity Instrument. | NR | NR<br>Development of the fidelity instrument in a different paper. |
| Miller, et al | 2017 | PT | PTCL cRCT | Fidelity will be measured through an audit of the fidelity checklist and electronic medical record (EMR) of each included patient to determine consistency of the intervention with the protocol | NR<br><br>Cites Hildebrand | consistency of the intervention with the protocol |
| Mitchell, et al | 2018 | SLP | FS RCT | Data about how the intervention was delivered (face to face or independently) and by whom were extracted from therapists' records. | NR | NR |
| Moore, et al, | 2018 | PA | SR, NS | n/a SR | Borrelli<br>Also TIDIER | 1)Treatment fidelity strategies for design of study.<br>Ensure same treatment dose within conditions<br>Ensure equivalent dose across conditions<br>Plan for implementation setbacks<br>2)Treatment fidelity strategies for monitoring and improving provider training<br>Standardise training<br>Ensure provider skill acquisition<br>Minimise "drift" in provider skills<br>Accommodate provider differences<br>3) Treatment fidelity strategies for monitoring and improving delivery of treatment<br>Control for provider differences<br>Reduce differences within treatment<br>Ensure adherence to treatment protocol<br>Minimise contamination between conditions<br>4) Treatment fidelity strategies for monitoring and improving receipt of treatment<br>Ensure participant comprehension<br>Ensure participant ability to use cognitive skills<br>Ensure participant ability to perform behavioural skills<br>5) Treatment fidelity strategies for monitoring and improving enactment of treatment skills<br>Ensure participant use of cognitive Skills<br>Ensure participant use of behavioural skills |
| Moran, et al | 2015 | PT<br>PA | FS | ? | Borrelli | ability of participants to comprehend both the visual and auditory components of the DVD. |
| Morrison, et al | 2017 | RH | FidSt | adherence to the training protocol, all intervention sessions were audiotaped. Samples from the audiotaped intervention sessions were reviewed by one of the principal investigators, who had developed a fidelity checklist to monitor intervention | NIH-BCC | 1) study design, (2) facilitator training, (3) intervention delivery, (4) intervention receipt, and (5) intervention enactment. |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | adherence. Facilitator trainers did in vivo observation of at least one session per facilitator. | | |
| Morse, et al | 2017 | PA | FS RCT? | Fidelity Checklist | NR | NR |
| Mosen g, Dagfinr ud, Østerås | 2019 | PT PA | RCT | Analysed study notes and self-reported questionnaire data. | Carrol | "Bottom-line measure of fidelity within this framework is the **evaluation of adherence.** Adherence should be addressed through the subcategories content and dose. To evaluate adherence in the SAMBA study, six core components representing content and dose of the implementation strategy were classified as either "high adherence", "partly adherence" or "low adherence" The components were: 1) Proportion GPs and PTs attending the workshops; 2) PT knowledge and attitudes after the workshop towards eight statements on evidence-based OA treatment; 3) Number of times the PT adjusted their patients' exercise programs; 4) Proportion of patients who received physiotherapy; 5) Proportion of patients who completed the patient education and exercise period and 6) Proportion of patients who exercised according to dose recommendations from ACSM." |
| Murph y, et al | 2011 | OT | PTCL | standardized modules for the intervention and OT and treatment receipt | NIH-BCC | Fidelity of Treatment Delivery: intervention groups are equivalent in dose and administration of treatment.<br>And fidelity of treatment receipt: participant involvement in treatment. |
| Murph y, et al | 2008 | OT | PILOT | NR | NR<br><br>Cites Bellg | NR<br><br>"Controlling for treatment time is considered necessary for treatment fidelity in intervention studies" |
| Murray ,et al<br><br>Murray et al | 2015<br><br>2019 | PT | RCT | Mentions importance of fidelity, and implementing the communication strategies more closely to protocol. | NR | NR |
| Mustia n, et al | 2017 | PA Onco | SR/MA | Treatment fidelity protocol as a variable. | NR | NR |
| Namasi vayam, et al | 2018 | SLP | PILOT | Providing therapy cues are crucial for monitoring treatment delivery fidelity | NR Cites Hinkley Borrelli | NR<br><br>Adherence to Rx protocol described. |
| Naylor, et al | 2006 | PA | FS/PP | assessed fidelity by comparing actual to prescribed amount of physical activity delivered, actual versus potential number of weeks logged and by examining the planned and implemented activities coded across the six zones | NR | NR<br><br>fidelity to the model |
| Neel, et al | 2019 | RH PT OT NR | PTCL RCT | highly manualized protocol is monitored and 10% of videos reviewed by an independent | NR<br><br>Cites Borrelli | NR<br><br>checklist of critical elements and adherence to protocol described. |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | reviewer. Standard care is also monitored using nursing logs in both groups. randomly sampled fidelity of treatment (FOT) measures will be collected. checklist of critical elements, self-scored. | | |
| Nielsen , et al | 2014 | PT | QS | Comprehensive training and mentoring by psychologists. Intervention Training workshop. | | |
| Noble, et al | 2018 | LI | FidSt | Checklist, process evaluation | NR Cites Dane | Fidelity includes exposure, adherence to content and quality of delivery |
| ***Norris, et al*** | ***2015*** | | | | (NIH) model | The US National Institute of Health (NIH) model of fidelity |
| O'Neal et al | 2018 | PR | FS | Fidelity of the PAI was assessed using the checklist published by Borrelli | Borrelli | |
| O'Shea, et al | 2016 | PT BChge | ScR | n/a | Bellg(NIH-BCC) | (1) design of study, (2) training providers (3) delivery of treatment (4) receipt of treatment (5) enactment of treatment skills |
| Ownsworth et al | 2013 | CR ML | PTCL RCT | monitored using a checklist based on Borelli's framework. Sessions will be audiotaped to enable therapists' adherence to the treatment protocol to be examined for a random sample (20%) of sessions by experts who are independent of the study | Borrelli | NR |
| Ownsworth, et al | 2017 | OT RH | RCT | Therapist adherence to the treatment protocol was examined for a random sample (15%) of audiotaped sessions 1 to 8 for each intervention using a checklist based on Borelli's framework. | Borrelli | Delivering core components of intervention according to the manualized treatment protocol |
| Ownsworth, et al | 2013 | OT RH | PTCL RCT | Therapist adherence to the treatment protocol examined for a random sample (15%) of audiotaped sessions 1 to 8 for each intervention using a | Borrelli | Delivering core components of intervention according to the manualized treatment protocol |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | checklist based on Borelli's framework. | | |
| Palmer, et al | 2015 | SLP | PTCL RCT | checklist guiding SLP selection of exercises based on the participant language profile identified during assessment. | NR | NR |
| Palsola et al, 2020 | 2020 | PA | FidStd | Semi-structured interviews of providers and participants- analysed with thematic analysis | Bellg/ NIH-BCC | Fidelity refers to the methodological strategies used to monitor and enhance the reliability and validity of behavioural interventions (Bellg et al,2004). However, interventions can be delivered with perfect fidelity, but not work as<br>intended if participants do not accept and engage with them.<br><br>Receipt refers to the participant side of fidelity, that is, comprehension of intervention and performance of the cognitive and behavioural skills taught in the intervention (Bellg et al, 2004) |
| Pastva et al | 2018 | PT RH | FidSt RCT | | NIH-BCC | 1. Ensuring the intervention dose is consistent across participants: Ensure that dose is adequately described and is the same for each participant within each condition and contamination across conditions must be minimized.<br>2 standardizing interventionist training. Ensure that training is conducted similarly across interventionists and maximize acquisition and maintenance of skills and to limit deviation from the standardized procedures over time.<br>3) monitoring intervention delivery. Ensure that intervention is being delivered as intended. Monitoring of intervention delivery involves assessment of intervention competency (did interventionists maintain the skill set learned in training), intervention differentiation (did the interventionists only deliver the target treatment and not other treatments), and intervention adherence (were intervention components delivered as intended).<br>4) evaluating participants' understanding of information provided. Ensure participant comprehends information to attend to and perform study-related skills. This particular aspect focuses on the participant's receipt of the study information, which is demonstrated by his/her ability to attend to and perform the physical rehabilitation exercises.<br>5) ensuring that participants use the skills taught in the intervention. Ensure the participant actually uses the skills provided in the intervention in appropriate life settings. |
| Patterson, et al | 2018 | SLP | Prosp. single cohort | content and treatment plans, recorded in patients' notes were evaluated by a CBT expert practitioner<br>as part of supervision | NR | fidelity were measured by assessing whether the intervention could be delivered as planned, by a SLT with CBT training |
| Pennington, et al | 2019 | SLP | PILOT RCT | Video recordings of 31 Skype dysarthria therapy sessions (19%) were checked for treatment fidelity. | NR | NR<br><br>adherence to the treatment protocol |
| Persch | 2013 | OT | PP | manual of procedures (MOP; operating procedures for the study and for training personnel in | NR | investigators should concern themselves with facets of treatment fidelity related to study design and the training of personnel. Fidelity practices related to study design help investigators discern whether the study will adequately achieve the aims and test the hypotheses that |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | the administration of outcome measures and interventions. Video Rx and regularly scheduled checks of both our outcome assessors and our intervention therapists | | have been set forth. Fidelity in training refers to the extent to which the outcome measures and treatment are administered in accord with the MOP. Thus, a team that has multiple protocol violations would be said to have low training fidelity, which would increase variability with which the protocol is administered. |
| Peters, et al | 2018 | HP | CRCT | Process evaluation | NR | NR<br>Fidelity gets heading but outcome not really described with data |
| Pfeiffer, et al | 2011 | OT | PILOT | Accordance with Parham SI Fidelity Measure | NR | NR |
| Poltawski, et al | 2014 | RH | PP | n/a | NHBCC | Study design Describing key ARNI elements and principles (A, B, C)<br>Ensuring intervention meets current best practice guidelines (C, D)<br>Developing intervention manual (A, B, C, D, F)<br>Identify appropriate assessment methods and outcome measures (A, E)<br>Identity process measures that might influence fidelity (A, B)<br>Provider training Identifying key elements of provider training regarding ARNI (B, C)<br>Identifying key elements of provider briefing regarding conduct of trial (A, C)<br>Developing trainer materials, quality standards and minimum experience levels (A, B, C, D)<br>Treatment delivery Distinguish core and flexible components of intervention (A, B, C)<br>Identifying necessary resources to deliver intervention (A, B)<br>Developing fidelity assessment instruments (A, F)<br>Identifying threats to fidelity and possible strategies to mitigate (A,B,F)<br>Treatment receipt Developing study participant information materials (A, C, E)<br>Identifying strategies to promote participant engagement (A, E)<br>Developing fidelity assessment instruments (A, B, F)<br>Treatment enactment Developing participant information materials (A, E)<br>Identify factors influencing adherence and ongoing engagement in intervention (A, B, E)<br>Developing fidelity assessment instruments (A, B, F) |
| Poltawski, et al | 2013 | PA<br>RH | CS<br>FS | participant interviews, audit of participant and EP records, and observation of training. | NR | NR |
| Pozehl, et al | 2010 | CR<br>PA | CT | | Bellg | Intervention fidelity strategies were based on the recommendations of Bellg et al The theoretic basis for specific intervention strategies was an important first step in assuring fidelity. Every participant in the treatment group received the same number of group sessions that were equal in frequency and length. The principal investigator led the group sessions for the treatment and control groups using a protocol developed from the Heart Failure Society of America's educational modules.26 The exercise training protocol was guided by a physical therapist in a cardiac rehabilitation setting during the first 3 weeks of the study. The principal investigator taught |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | resistance training during the first 3 weeks of the study using detailed pictures and guidelines. Receipt of the treatment was assessed during each group session through questioning and verification of participant understanding. The principal investigator or physical therapist observed enactment of skills on a regular basis during the first 3 weeks of the study and every 4 weeks during the remaining 9 weeks. |
| Pyatak, et al | 2017 | OT<br>LI | PILOT RCT | therapists document their adherence to the intervention protocol in treatment notes for each intervention session. Second, approximately 10% of sessions are observed by another therapist trained in the intervention, who completes a fidelity checklist and provides feedback to the treating therapist. Finally, weekly meetings are held with the full intervention team to facilitate problem-solving and prevent intervention drift. | NR | NR<br><br>intervention adherence, protocol deviations |
| Quinn, et al | 2016 | PT<br>PA | RCT | combination of self-report checklists, independent analysis of audio-recordings, and a self-assessment completed by the intervention coaches. | NR (Social grounded theory?) | |
| Reddington, et al | 2017 | PT | PTCL MMS | Implementation fidelity testing will be carried out in order to assess the treating clinicians are delivering what is intended by the protocol. An independent assessor will review video footage of physiotherapist and participant session in order to assess implementation fidelity. | MRC NIH-BCC | treating clinicians are delivering what is intended by the protocol |
| Reeves, et al | 2017 | PT<br>ExPhy | Pilot RCT | in-person training of the study interventionists. Ongoing oversite of study rehabilitation sessions. Bi-weekly intervention teleconferences among all site intervention leaders and interventionists | NR | NR |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | provide continued monitoring and guidance. | | |
| Resnick, et al | 2011 | RH EP | FidStPP | quantification of the sessions attended and activities that occurred within those sessions. delivery of the intervention was qualitatively evaluated based on 20-random observations using a checklist that included both control and treatment group interventions | NIH-BCC Moncher | when designing a study so as to maximize treatment fidelity, three issues need to be considered: treatment delivery, receipt, and enactment. Delivery focuses on assuring that the intervention was delivered as proposed to all participants. Receipt addresses whether the participant understood the intervention, learned new information, and can perform a new behaviour and therefore expands beyond just exposure or delivery of the intervention to the individual. Enactment seeks to assure that the intervention is performed in real world settings. Treatment fidelity for an intervention should also be evaluated with regard to study design and training of interventionists. Treatment fidelity related to design considers adherence to group assignment and explores the degree to which the theoretical framework on which the intervention was developed is maintained. Design: Adherence to the underlying theory related to exercise and prior work testing similar interventions. Assurance that treatment group was exposed to the treatment intervention and control group exposed to the control intervention and that there was no carryover between groups. Treatment fidelity related to the training of interventionists assures that those implementing the intervention were adequately prepared to do so and implemented all aspects of the treatment as intended. |
| Rich, et al | 2017 | PA | ImpSt | qualitative and quantitative measures tracked over time by peer health coaches with tracking tablet to assesses and maintain the quantity and fidelity of the program delivered. Interviews. | NR | fidelity (quantity & quality of intervention delivered) |
| Robbins, et al | 2016 | PA | FidStd | Process evaluation, survey to reflect extent to which intervention reflects conceptual framework | | Measurement of “fidelity”, which can also be accomplished via survey, assists in determining the extent to which the intervention is consistent with the conceptual framework on which it is based. extent to which intervention reflects conceptual framework (DOSE NOT PART OF FIDELITY HERE). Fidelity: theoretical integrity |
| Robinson et al | 2021 | Surg | SR | Evaluates reporting of intervention adherence in surgical trials. | NR | Intervention adherence reporting evaluated in the systematic review, but not defined and no theoretical model/framework cited. |
| Roberts, et al | 2018 | RH | PrEv Randomised FS | Process evaluation | NR | NR<br><br>Measurement of intervention fidelity: completion of workbook tasks, completion of diaries and number and content of therapy sessions.<br>Delivery/ delivery to individuals. What rehabilitation intervention is delivered? Is it what was intended by the researchers? What intervention is delivered to each participant? Is the delivered intervention the one intended by the researchers?<br>Theory: What theory has been used to develop the intervention?<br>Context: What is the wider context in which the feasibility study is conducted? |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | Response of rehabilitation teams/patients? How is the enhanced intervention adopted by the rehabilitation teams? How do the patient participants respond? |
| Robins, et al | 2019 | PA LI/HP | PP CTrial | Manual, provider training, multimedia materials. Observation of Rx and post-Rx interviews. | NIH-BCC | 1) treatment design, (2) training providers, (3) delivery of treatment, (4) receipt of treatment, and (5) enactment of treatment skills. design should ensure that the study hypotheses are tested in relation to both underlying theory and clinical processes. Provider training is focused on assessing and optimizing training processes for intervention delivery. Ensuring fidelity through treatment delivery is focused on processes that ensure the treatment is delivered as designed and focus on standardizing and improving delivery as well as assessing adherence. Processes of treatment receipt involve monitoring and optimizing participant understanding and performance of intervention skills during treatment delivery. treatment enactment focus on ensuring that cognitive and behavioural intervention elements are applied in relevant daily life situations. |
| Rodgers, et al | 2017 | RH | PTCL RCT | TIDIER items 10-12 Data from robot software and training sessions are periodically reviewed to monitor intervention adherence and feedback is provided to therapy staff delivering the intervention | NR | How well (planned)? 'If intervention adherence or fidelity was assessed, describe how and by whom, and if any strategies were used to maintain or improve fidelity, describe them. How well (actual)? 'If intervention adherence or fidelity was assessed, describe the extent to which the intervention was delivered as planned' |
| Ryan et al | 2016 | PT | PTCL RCT | Fidelity of the resistance training programme to trial protocol will be quantified by observations of exercise sessions. Semi-structured interviews. Process eval. | NR | NR<br>Fidelity of the resistance training programme to trial protocol. |
| Salamh, et al | 2019 | PT PA | SR/MA | Modified NIH-BCC checklist to create customized fidelity Checklist for the systematic review/Meta-Analysis. | Borrelli (NIH-BCC) | "Item 1: Was information about the treatment dose in the intervention condition provided?<br>Item 2: Was information about the treatment dose in the control or comparison condition provided? Item 3: If more than 1 intervention was described, were they all described equally well?<br>Item 4: Were methods used to ensure the dose was equivalent between conditions?<br>Item 5: Were methods used to ensure the dose was equivalent within a condition?<br>Item 6: Were characteristics to be sought and avoided by the treatment provider addressed *a priori*, and was some mention made of credentials? Item 7: Was there a mention of a theoretical model or clinical guidelines on which the intervention was based?<br>Item 8: Did the authors indicate how providers were trained? Did the authors indicate that provider training was standardized? Item 9: Was there a method to ensure that the content of the intervention was being delivered as specified?<br>Item 10: Was there a method to ensure that the dose of the intervention was being delivered as specified?<br>Item 13: Were nonspecific treatment effects evaluated?" |

| | | | | | | |
|---|---|---|---|---|---|---|
| Salamh , et al | 2015 | PT | SR/MA | Modified NIH-BCC checklist to create customized fidelity Checklist for the systematic review/Meta-Analysis. | Borrelli (NIH-BCC) | "Item 1: Was information about the treatment dose in the intervention condition provided?<br>Item 2: Was information about the treatment dose in the control or comparison condition provided? Item 3: If more than 1 intervention was described, were they all described equally well?<br>Item 4: Were methods used to ensure the dose was equivalent between conditions?<br>Item 5: Were methods used to ensure the dose was equivalent within a condition?<br>Item 6: Were characteristics to be sought and avoided by the treatment provider addressed *a priori*, and was some mention made of credentials? Item 7: Was there a mention of a theoretical model or clinical guidelines on which the intervention was based?<br>Item 8: Did the authors indicate how providers were trained? Did the authors indicate that provider training was standardized? Item 9: Was there a method to ensure that the content of the intervention was being delivered as specified?<br>Item 10: Was there a method to ensure that the dose of the intervention was being delivered as specified?<br>Item 13: Were nonspecific treatment effects evaluated?" |
| Salmoir ago Blotche r | 2017 | PA CR | UCT | sessions were video recorded, and the study auditor reviewed 10% of videos for protocol consistency using a checklist developed for the study. | NIH-BCC | Not reported. |
| Sandbo rgh,et al<br><br>Part I and Part II | 2010 | PT | FidSt | PT training in intervention and how to adapt/tailor. Rx components were operationalized in accordance with the treatment manual. Integrity checklist developed for study: documentation, treatment components categorized as present or not. Treatment documents reviewed/ evaluated by second author not involved in training of PTs, or in supervision intervention phase. | NR<br><br>Cites Pereple tchikov a | Treatment integrity includes three components: adherence, competence, and differentiation). Adherence is the degree to which the therapist conducted treatment in adherence with the treatment manual; overall and for treatment components. Competence refers to the therapists' level of skill when delivering the treatment. The therapists' degree of competence could account for a potentially large degree of variance not explained by the treatment itself and may vary from patient to patient and depend on many extraneous factors. Differentiation refers to whether treatments in an intervention differ from each other in the intended manner and is closely related to therapist treatment adherence, i.e. that treatments are distinctly different from one another and do not overlap. |
| Schaaf, et al | 2012 | OT | FidSt | Manualisation of the intervention and examination of the treatment manual's adherence to fidelity Random sampling of 20% of the videotapes of the treatment sessions were rated for therapist's fidelity | NR<br><br>Cites Bellg | NR<br><br>fidelity to the manualized intervention. |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | with sensory fidelity checklist. | | |
| Schaaf, et al | 2014 | OT | RCT | Manualisation of the intervention and examination of the treatment manual's adherence to fidelity Random sampling of 20% of the videotapes of the treatment sessions were rated for therapist's fidelity with sensory fidelity checklist. | NR<br><br>Cites Bellg | NR<br><br>fidelity to the manualized intervention. |
| Schepens, et al | 2012 | OT<br>PA | RCT | OT protocol training module with the PI/written session reports regular check-in phone meetings with PI. | NIH-BCC | NR |
| Scott, et al | 2018 | PA | FS | Three 10-min segments were analysed from separate audio-tapes by an independent coder using checklist/tool. Provider training before RX to ensure competence/delivery. | NR | Treatment fidelity was assessed, including compliance to intervention delivery. protocol deviations. Assessing fidelity of intervention delivery can optimize intervention effectiveness by identifying and correcting protocol deviations early and help sustain practitioner's skills |
| Sharma, et al | 2019 | CP | RCT | NR | NR | Not eligible- fidelity not assessed, though in protocol |
| Shivonen, et al | 2013 | Surg | RCT | All procedures were standardized and recorded on video | NR | NR<br><br>Adherence to study protocol. |
| Shrubsole, et al | 2018 | SLP | Pilot cRCT | Intervention delivery checklist (self-reported) | MRC guidance on feasibility. | Checklist has "core information about each intervention session (such as date, duration and number of participants) and the extent to which the components of each intervention was delivered as planned (e.g., PowerPoint presentation, video of person with aphasia)" (see Appendix D for details) |
| Silveira, et al | 2019 | PA<br>RH | FidSt PTCL | Fidelity monitoring according to NIH-BCC guidance. | NIH-BCC | Study design: focuses on practices that ensure study procedures and implementation are in line with current theory and clinical processes. Study design fidelity goals include ensuring that conditions are congruent with relevant theory and practice, ensuring equivalent treatment dose within and across conditions, and planning for implementation setbacks.<br>Provider training: address preparation for uniform delivery of treatment.<br>delivery of treatment: focuses on ensuring the intervention is delivered as intended. Many of the concerns within delivery of treatment overlap with strategies for training and study design, including controlling for provider differences and adhering to created protocols; however, this area further addresses differences within treatment conditions and minimizes contamination receipt of treatment: involves strategies and monitoring of a participant's ability to understand and adopt treatment-related behavioural skills and cognitive strategies.<br>enactment of treatment: strategies aimed at monitoring and improving participant ability to perform treatment- |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | related behavioural skills and cognitive strategies in relevant real-world settings. |
| Skidmore, et al | 2014 | RH<br>Stroke | PS<br>PE | All research intervention sessions were videotaped and rated for fidelity to the respective manualized procedures with validated, tailored fidelity checklist. | NR<br><br>Cites Hildebrand | We examined two facets of fidelity: 1) treatment integrity, and 2) treatment differentiation. To address treatment integrity, independent raters trained in the respective protocols assessed adherence to specified principles in each protocol (yes, no), and competence in execution (inadequate, adequate, exceptional). To address treatment differentiation between the two protocols, raters assessed adherence of both research interventions to the strategy training protocol to determine the degree to which the strategy training sessions adhered to the planned protocol, and the degree to which the attention control sessions did not include elements of the strategy training protocol. |
| Skolasky, et al | 2013 | PT | CT | Audio recordings are made for all telephone calls with all participants to assess the fidelity of the intervention with MI tool/checklist. ongoing monitoring of intervention integrity. Monthly booster sessions. | NR | NR<br><br>Fidelity not defined or reported. |
| Smith, et al | 2019 | PT<br>PA | FS<br>MMS | analysis of PTs clinical notes against a three-point checklist outlining important details and components of intervention to be completed by the PT. The three-point checklist included: specific pain education; delivery of a loaded exercise programme; and discussion on self-management strategies, | NR | NR<br><br>Fidelity was defined as adherent and competent delivery of the intervention, |
| Söderlund, et al | 2009 | PT<br>Pain | PTCL, RCT | detailed treatment manual and a treatment protocol/ checklist manual has also been developed to guarantee that the treatment will be unchanged during the course of the study. | Bellg | To ensure the same treatment within condition a detailed treatment manual and a treatment protocol/ checklist is used for each patient separately manual has also been developed to guarantee that the treatment will be unchanged during the course of the study. Both IT-based and Face-to-face-based interventions are going to have equal number of treatment sessions/phases to ensure equivalent dose across conditions. 4. One therapist will deliver both group treatments to ensure standardized trained therapist and to minimize contamination between conditions.<br><br>5. Patients can e-mail their questions in IT-group to the therapist. These questions, we believe, will mirror patient's understanding of the treatment. Also, often asked questions and answers will be on the IT-group's home page for all patients in this group to read.6. By asking so called consumer questions, we are recording patients' beliefs and expectations about the intervention and also, if the expectations are fulfilled.7. To ensure that the patients are able to use cognitive and behavioural |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | skills we are applying home exercises in daily activities. These are reported in a diary and always dis- cussed with the patient.8. To ensure that the behavioural components are not given to the standard care-group in the acute stage the therapist follows a strict manual only dealing with physical symptoms and advice given for all patients at the initial visit. 9. The number of intervention contacts (e-mail contacts and Face-to-face group meetings) is reported. |
| Sosnowski et al | 2018 | RH | RCT<br>FS | retrospective recording of care data on a case report form. | NR | Successful adherence to the protocol was defined as the administration of the entire prescribed ABCDE bundle on at least 80% of ventilated days. |
| Sprowson, et al | 2014 | Surg | PTCL<br>RCT | n/a: only effect of options for randomization on RX fidelity discussed. | NR | NR |
| Stephens, et al | 2018 | Surg | PE<br>cRCT | 37-item, online questionnaire, administered at the end of the study period. Sample of interviews audio recorded, and field notes recorded in a diary at the time of observation, or immediately afterwards. | | design of the intervention and the operational elements required for effective delivery. design (or programme) level and the hospital (operational) level. At the design level, adaptability is often essential in ensuring that quality improvement interventions can fit within different contexts. fidelity to key parts of an intervention is also important to maximise likelihood of success |
| Stephens, et al | 2018 | PS | RCT | routine QI programme activity data (records of meeting attendance and use of the web-based resource) data from an exit questionnaire sent to all QI leads and ethnographic data. The 37-item, online questionnaire, administered at the end of the study period. | Carroll<br><br>Referenced but NR | NR<br><br>Adaptation/fidelity results discussed. |
| Stevens, et al | 2007 | RH<br>PT<br>OT<br>SLP | PP<br>ImpSt | To ensure consistency and accuracy in the delivery of the workshop, we developed information feedback, individual consultations, and detailed outlines of all intervention components and established a timeline to ensure timely delivery of all intervention activities. | NIH-BCC<br>TI | The goal was to achieve a match between the written protocol of the intervention and the research staff's actual delivery of the treatment. Treatment Implementation (TI) framework: ability to deliver the intervention according to a specific and predefined protocol and that participants perceive and understand the treatment as intended. Participants' ability to enact the skills or behaviours outside the intervention or training setting. |
| Strasser, et al | 2008 | RH<br>PT<br>OT<br>SLP | cRCT | To ensure consistency and accuracy in the delivery of the | NIH-BCC<br>TI | The goal was to achieve a match between the written protocol of the intervention and the research staff's actual |

| | | | | | | |
|---|---|---|---|---|---|---|
| (follows Stevens, et al) | | | | workshop, we developed information feedback, individual consultations, and detailed outlines of all intervention components and established a timeline to ensure timely delivery of all intervention activities. | | delivery of the treatment. Treatment Implementation (TI) framework: ability to deliver the intervention according to a specific and predefined protocol and that participants perceive and understand the treatment as intended. Participants' ability to enact the skills or behaviours outside the intervention or training setting. |
| Stuart, et al | 2009 | PT | FS | PT observation of classes to ensure adherence. | | Fidelity: exercise protocols are being followed. |
| Sturkenboom, et al | 2013 | OT | FS RCT | analysis of protocol adherence. Assessors used an assessment log to register duration of the visit, adherence to the assessment protocol and any irregularities encountered. | NR | NR<br><br>adherence to the protocol and actual treatment delivery (process, content, and time) |
| Sturkenboom, et al | 2016 | OT | PE | Process evaluation with analysis of protocol adherence. | Gearing | The treatment fidelity: the dose, the protocol process adherence and content of treatment delivered compared to the protocol. 'treatment fidelity', which is defined as the extent to which the intended intervention was provided by the therapists.<br><br>NOT PART of FIDELITY: Another concept is 'treatment enactment', the extent to which recipients (i.e. patients and caregivers) apply the interventions in daily life. |
| Swank, et al | 2003 | Surg | RCT | Surgeons were allowed to apply their own techniques within limitations of the protocol. Procedures and adhesion assessments were recorded on video, and outcomes were reviewed by two surgeons. | NR | NR<br><br>Adherence to protocol. |
| Tang, et al | 2018 | PA | SR | n/a | Carrol, Frank Tomila and Braun | **Carroll**: Conceptual Framework for Implementation Fidelity:<br>Adherence: Content, coverage, frequency, duration.<br>Moderators: Intervention complexity, Facilitation Strategies, Quality of delivery, Participant responsiveness.<br>Identify: essential elements.<br>**Frank: (NIH-BCC)** 1. Intervention design, 2. provider training 3. treatment delivery of 4. receipt of treatment 5. treatment enactment.<br>**Tomika:** Four Step Fidelity Assurance Protocol:<br>1.Deconstruct program and prepare implementation plan.<br>2.sponser staff training.3.Monitor using standard checklists. 4.Track participant outcomes. |
| Tarrant, et al | 2018 | SLP | PTCL Pilot RCT | Session check lists, completed by facilitators, will capture whether the main content of the | NR | NR<br><br>For analysis of intervention fidelity and engagement we will use trainer interview data, session checklists completed by facilitators (that will be part of the |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | Intervention Manual is being delivered, indicate where flexibility of delivery is permitted (in session structure/content) and allow evaluation of intervention fidelity. Intervention fidelity will be assessed by several methods: singing group attendance, session checklists, observations, and video recordings of selected singing group sessions in the programme. *Researchers assess this.* | | Intervention Delivery Manual), participant attendance data, researcher observations and videos of singing sessions. |
| Taylor, et al | 2015 | RH | PTCL RCT | process evaluation will assess fidelity of intervention delivery. A fidelity checklist developed as part of the programme will be used to assess fidelity of delivery of the intended intervention processes. This will be achieved by analysing recordings of all contacts (telephone and face to face) between intervention facilitators and 20 purposively sampled patient participants. | MRC | how well (or otherwise) intervention components are delivered and received and will also allow researchers to describe variability in fidelity of delivery across patients and facilitators. |
| Taylor, et al | 2015 | PA CR | FidStd | heart rate data recorded for these participants (n = 17; 7 females) to illustrate our fidelity assessment method | Resnick | Intervention fidelity refers to the extent an experimental manipulation has been implemented as intended in a comparable manner to all participants. address session attendance and compliance (meeting the prescribed exercise intensity), as this interaction constitutes the dose of the intervention and influences the physiological response to exercise training. to quantify the overall dose of the intervention, intention to treat fidelity analysis should include all participants irrespective of their attendance and compliance. per protocol fidelity analysis should involve only those participants who attended all of the<br>prescribed sessions. Both approaches are informative in a full exploration of fidelity. |
| Tew, et al | 2016 | PA | SR | TIDIER checklist | NR | NR<br>intervention adherence or fidelity |
| Thakur, et al | 2012 | Surg | ER | Fidelity to intended target. | NR | NR |
| Thomas, et al | 2018 | SLP | UCT FidST | PI assessed parent and clinician treatment fidelity and reliability of perceptual | NR<br><br>Cites Kaderavek | We can divide the elements of fidelity into perceptual and procedural components. The perceptual component of fidelity is evaluated through measurement of reliability of perceptual judgements of the child's speech. In contrast, |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | judgements on randomly selected 10 minutes of the practice phase of each clinician-delivered session, and 100% of each parent-delivered home-based session. | | fidelity for procedural aspects of the treatment, such as giving feedback after a 3–7second delay. |
| Thompson,et al | 2018 | PA LI | FidSt | Recordings of sessions scored with checklist by authors. | NIH-BCC | Study Design, Provider Training, Treatment Delivery, Treatment Receipt, and Treatment Enactment |
| Toomey, et al | 2016 | PT | FidSt | direct observations, audio recordings, and self-report checklists. The direct observations were conducted using a checklist developed by the research team to assess the fidelity of the delivery of sessions and the treatment dose. | NIH-BCC | Study design-addresses factors that should be considered when designing the trial and are intended to enable the study to adequately assess its hypotheses in relation to the underlying theory and mechanisms of action of the study.<br>Training of providers-aims to ensure and assess that providers are able to deliver the intervention satisfactorily and as intended<br>Treatment delivery -relates to processes that assess and enhance the actual delivery of the intervention so that it is delivered as intended<br>Treatment receipt- involves using strategies to enhance and assess participant knowledge and use of intervention skills and learning during the intervention. It also considers factors that aim to enhance the acceptability of the intervention to the participant<br>Treatment enactment- uses strategies to enhance and assess their actual practice of the intervention skills and knowledge in daily life |
| Toomey, et al | 2015 | PT | RP | NIH-BCC checklist | NIH-BCC | Fidelity practices related to Study Design are factors that should be considered when designing the trial, and are intended to enable the study to adequately assess its hypotheses in relation to the underlying theory and mechanisms of action of the study (e.g. establishing the behaviour change theory underpinning the intervention and outlining proposed methods to assess its implementation), whilst Training of Providers assesses and ensures that the providers *can* deliver the intervention satisfactorily (e.g. procedures put in place to train the providers, and also procedures to assess the effectiveness of this training). The domain of Treatment Delivery relates to the monitoring of actual intervention delivery (e.g. direct observation of intervention sessions to evaluate delivery of the behaviour change theory techniques) whereas Treatment Receipt and Enactment both focus on the recipient of the intervention, or the participant; using strategies to enhance and monitor participant knowledge and use of intervention skills during the intervention (Receipt) and using strategies to enhance and monitor their actual practice of the intervention skills and knowledge in daily life (Enactment). |
| Tuntland, et al | 2015 | OT RH | RCT | Not adequately monitored. | NR Cites Bellg | NR<br><br>Treatment fidelity, i.e. if the treatment was delivered as intended |
| Tully, et al | 209 | PA | PILOT RCT<br><br>PE | Training/support manual. structured observation of intervention delivery by a member of the research team responsible for | NR | NR:<br><br>Fidelity: Fidelity of delivery and receipt of intervention |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | mentor training, semi-structured interviews and focus groups with peer mentors and participants as part of the post-intervention follow-up. audio-recorded sessions to assess the content fidelity of delivery. Fidelity checklists. | | |
| Tyson, et al | 2015 | RH PT | RCT | Not assessed | Hennessey ** | fidelity to the treatment protocol. |
| van Bysterveldt | 2010 | SLP | | | | |
| Vaughan-Graham, et al | 2014 | RH PT OT | SR | | Hildenrand | This refers to the extent of standardization of the actual intervention, as well as to the details on who provides the intervention including how they are trained and supervised throughout the study. Intervention fidelity also requires that the appropriate background and experience level of the study therapists is identified and ensured<br>The ability to operationalize and standardize the intervention, as well as quantify the level of skill of the therapist, supervise and evaluate adherence of the intervention<br>The degree to which a therapist implements an intervention under research conditions (treatment fidelity) is dependent upon the extent and operationalization of the intervention and skill level of the therapist.<br>level of training, skill, or evaluation of adherence of the therapists |
| Volkmer, et al | 2018 | SLP | PTCL RCT | Video/audio recording assessment, recording of sessions and analysis of random selection with treatment adherence checklist. Ind.Raters | NR | NR<br><br>measures of fidelity will demonstrate the consistency with which the intervention in delivered.<br><br>an assessment of treatment fidelity to determine necessary levels of SLT training,<br><br>fidelity measures (video recordings, local collaborator adherence questionnaire, and participant feedback questionnaires) |
| von Thiele Schwarz, et al | 2015 | LI | FidStPP | Fidelity checklist from adapted CFIF | CFIF | Conceptual Framework for Implementation Fidelity. three aspects of fidelity<br>in the framework (content, coverage, and dose) were complemented with a fourth aspect, namely timeliness: i.e. if the intervention is carried out at the right time.<br>fidelity involves assessing adherence, including its subcategories - content, frequency, duration (dose), and coverage. Thus, adherence relates<br>to whether participants have received the active components of the intervention as often and for as long as initially planned. It also relates to whether all the individuals who should be participating or receiving the benefits of an intervention are reached. |
| Vranceanu, et al | 2019 | Pain | PS FS | Adherence checklist | NR | NR |
| Walker, et al | 2016 | RH LI/HP | FS RCT | 'train the trainer' manual, regular | NR | strict protocol compliance |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | supervision, and checklist monitoring. | | |
| Wang, et al | 2014 | PT | PILOT RCT | PT monitoring of exercise performance and training intensity to ensure the treatment fidelity | NR | NR<br><br>Fidelity not defined or reported. |
| Watling, et al | 2007 | OT | SST | rated videotapes of random selection of sessions (21%) with SI tool. | NR | NR |
| Watson, et al | 2017 | PA | SR | | | |
| Wells, et al | 2016 | RH SLP | PTCL | Manual | NR | NR |
| Wenborn, et al | 2016 | OT | PTCL RCT | audio record COTiD-UK sessions and transcribe a sample to monitor the occupational therapists' adherence to the intervention, using a checklist derived from the original study. | NIH-BCC | programme adheres to the intervention manual cover five domains: study design, provider training, intervention delivery, intervention receipt, and intervention enactment |
| Wesson, et al | 2013 | PT OT | FS Pilot RCT | Adherence to the intervention protocol was recorded using field notes during each visit and included comments regarding acceptability of study components – whether participants were engaged in the exercise and/or home safety interventions. | | NR<br>Adherence to the intervention protocol.<br>acceptability of study components – whether participants were engaged in the exercise and/or home safety interventions. |
| Westland et al | 2017 | PA LI | PTCL cRCT | nurses allocated to the study arm will randomly audiotape one consultation from among the four consultations. The audiotapes will be coded using a coding list developed specifically for this study, consisting of the content of each of the four consultations and the Behaviour Change Counselling Index | NR<br><br>Cites Bellg | NR |
| Weston, et al | 2017 | PA CR Surg | UCT | detailed evaluation of the exercise sessions | Resnick | Intervention fidelity refers to the extent an experimental manipulation has been implemented as intended in a comparable manner to all participants. address session attendance and compliance (meeting the prescribed exercise intensity), as this interaction constitutes the dose of the intervention and influences the physiological response to exercise training. An assessment of fidelity permits an |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | Understanding of whether the exercise was performed at the Prescribed intensities, at all study sites and throughout all phases of the study. |
| White, et al | 2019 | PR | FS CRCT | Face to face, phone, smart phone recordings transcribed verbatim to facilitate assessment of intervention fidelity. | NIH-BCC | NR |
| Whitney, et al | 2013 | OT | NR | n/a | Gearing | Intervention fidelity examines the extent to which the intervention is delivered as it was intended. Describing the specific intervention protocol, which might be in the form of a manual, is the first step toward being able to achieve fidelity in adherence to the protocol |
| Wilbur, et al | 2016 | PA | FidSt | Breitenstein's Fidelity Checklist and fidelity manual. digital audio recording assessment, Enactment was measured by assessing participants' self-monitoring of their lifestyle PA prescription | NIH-BCC | fidelity delivery, receipt, and enactment. study design assures that treatment effects are not confounded with extraneous differences between the treatment and control condition. Treatment fidelity related to training of interventionists assures that interventionists are satisfactorily trained to deliver the intervention to the participants. Treatment fidelity related to delivery of treatment considers that the interventionist delivers the intervention as intended. Treatment fidelity of receipt of treatment focuses on exposure of the participant to the intervention and their ability to understand the skills and perform the treatment-related behaviour skills during treatment delivery. enactment is how well the participant can apply the treatment-related behaviour skills. |
| Williamson, et al | 2018 | PT | RCT | structured record of the interventions (treatment log is completed by the physiotherapists and used to monitor fidelity. | NR<br><br>Cites Bellg | Adherence with the intervention (attendance and the participants' engagement with the programme rated by the physiotherapist). |
| Williams, et al | 2015 | PA | cRCT | SEE FRENCH 2011 | Bellg | SEE FRENCH 2011 |
| Wilson, et al | 2009 | PA | PE | Evaluation against essential elements framework. data collected by a trained, *independent process evaluator* using systematic observation of after-school program activities, checklist. To assess dose and fidelity, the process evaluator observed sessions. | essential elements framework,<br><br>SDT* | Fidelity: fidelity and dose (completeness) of implementation. essential elements informed the development of dose (completeness) and fidelity. essential elements framework that defined dose and fidelity or "complete and acceptable delivery" of the ACT intervention. 1) Fidelity (for PA and behavioural skills components)- To what extent was the social environment autonomy supportive? 2) Dose delivered (completeness for all components)-To what extent were all planned components of the program provided to program participants? and 3) Reach-What percentage of the possible target group attends each week of the program? |
| Wilson, et al | 2010 | PA | IMPST | participants given a manual. ite co-ordinator and project director reports | Durlak | Fidelity: fidelity (degree to which the protocol was implemented as planned), the extent to which the intervention has been received by the audience.<br>defining the active ingredients: using theory or past research to delineate the intervention's active ingredients in clear operational terms, which should be guided by beliefs explaining why they should be successful<br>(ii) using good methods to measure implementation: developing an accurate and valid system for assessing implementation. This should include assessing both the fidelity and dose. DOSE NOT PART OF FIDELITY HERE.<br>(iii) monitoring implementation: assessing the program's active ingredients throughout implementation |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | (iv) relating implementation to outcomes: using implementation data to better understand program effects |
| Winstein, et al | 2013 | OT RH | PTCL RCT | Expert reviewer assesses the digital video footage accompanying documentation for therapist mastery of each ASAP principle implemented during a 1 hr session. Ongoing training/supervision. | NR<br><br>Cites Bellg | NR<br><br>execution of intervention adherence |
| Wong, et al | 2019 | OT RH | FS | Delphi consensus | Cross ** Hildebrand | Fidelity' generally refers to the degree to which a programme is delivered as intended by its developers. However, simple adherence to the manual is only part of the picture; there is an increasing call for a distinction between adherence and competence12,13 in measuring treatment fidelity. |
| Woolf, et al | 2016 | SLP | FS qRCT | fidelity checklist was developed, which covered each stage of the treatment protocol as described in the manual and recorded any deviations from it. Video of Rx. *Independent assessment.* | NR | The current study coded individual therapist behaviours as compliant/not compliant with the treatment manual. This very stringent procedure showed that there were deviations from the protocol, e.g. because cues prescribed in the manual were omitted or augmented. |
| Wright, et al | 2019 | SURG | | The proportion of management protocol components completed as intended will be assessed using a checklist at the time of preformed silo application and defect closure. The checklist will be completed by the person undertaking the intervention for every neonate included in the study. A second observer, who has been trained in the gastroschisis management protocol, will independently complete the checklist for 50% of the cases. | Cites Schoenwald and Cohen | Fidelity: The proportion of management protocol components completed as intended. protocol uptake and fidelity. compliance with the protocol (fidelity). important to distinguish between non-compliance and purposive adaptations<br><br>Cohen DJ, Crabtree BF, Etz RS, et al: Fidelity versus flexibility: translating evidence-based research into practice. Am J Prev Med. 2008; 35(5 Suppl): S381–9. Schoenwald SK: It's a Bird, It's A Plane, It's … Fidelity Measurement In the Real World. Clin Psychol (New York). 2011; 18(2): 142–7. |
| Yates, et al | 2013 | CR PA | FidStPP | Checklists to assess congruence with delivery of the components of CR, random observations. study manual was created | NIH-BCC | study design, training providers, delivery, receipt, and enactment of intervention skills. Design: delivery consistent within and across CR clinical sites so that a study can adequately test its hypotheses in relation to the underlying theory and clinical processes. Training: assessment and ongoing evaluation of the training of interventionists. Delivery: intervention is delivered as intended. It refers mainly to |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | actions of the interventionist. Receipt: treatment has been received and understood. Enactment: skills used in real-life settings as intended. fidelity components of design, training, and delivery of the intervention were the most different from fidelity in typical intervention studies. |
| Yu-Yahiro, et al | 2009 | PA | RCT | Treatment fidelity visits were performed by 2 investigators on 5 different exercise trainers. Trainers were observed a total of 70 times | NR | study design, training providers, delivery of treatment, receipt of treatment, and enactment of treatment skills |
| Zingmark, et al | 2014 | OT | RCT | Participant attendance rate measured, supervision and interviews with therapists. | Borelli/ NIH-BCC | Fidelity measured but never defined or described. E.g.," From our study, the possibility to draw conclusions about feasibility is limited to attendance rate and programme fidelity." |

**Study design:**

| | | |
|---|---|---|
| CP: Consensus paper | CT: Controlled study (non-randomised) | DS: Development study |
| ER: Experimental Results | FidSt: Fidelity study | FS: Feasibility Study |
| ImpSt: Implementation Study | PS: Pilot Study | PTCL: protocol |
| PE: Process Evaluation PP: | PS: post or pre surgical care. | Perspective or methodological paper |
| RCT: Randm. Controlled Trial | RP: Review paper | SR: Systematic Review |
| UCT: Uncontrolled clinical trial | QS: Qualitative Study | |

**Field/discipline:**

| | | |
|---|---|---|
| CR: Cardiac rehab | Complex Interventions | CRML: Cognitive Rehab-Motor Learning |
| CBT: Cognitive-Behav.Therapy | EP: Exercise Physiology/ists | LI/HP: Lifestyle/General health, health promotion (with physical component) |
| OT: occupational Therapy | PA: Physical Activity/Exercise | PT: Physical Therapy |
| PR: Pulmonary Rehab | RH: Rehabilitation-mixed | SLP: Speech Therapy |

**Other:** "*" Paper may contain more detail. n/a: not applicable NR: not reported

Appendix III: Chapter 3 supplementary material

Table 3.1: Participant checklist responses, raw data

| Partipant | Study | Item 1 | Item 2 | Item 3 | Item 4 | Item 5 | Item 6 | Item 7 | Item 8 | Q1 Fid. Supp | Q2 Fid. Mon | Q3 Fid Rep. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 2 | 2 | 1 | 2 | 2 | 3 | 0 | 3 | 2 | 2 | 3 |
| 2 | 1 | 2 | 2 | 2 | 2 | 1 | 3 | 0 | 3 | 2 | 2 | 3 |
| 3 | 1 | 2 | 2 | 0 | 2 | 2 | 3 | 1 | 1 | 2 | 2 | 1 |
| 4 | 1 | 2 | 2 | 1 | 2 | 2 | 3 | 0 | 3 | 2 | 2 | 3 |
| 5 | 1 | 2 | 2 | 1 | 2 | 2 | 3 | 0 | 3 | 2 | 2 | 3 |
| 6 | 1 | 2 | 2 | 1 | 2 | 2 | 2 | 0 | 3 | 2 | 2 | 3 |
| 7 | 1 | 2 | 2 | 1 | 1 | 2 | 3 | 0 | 3 | 2 | 2 | 3 |
| 8 | 1 | 2 | 2 | 1 | 2 | 2 | 3 | 0 | 3 | 2 | 2 | 3 |
| 9 | 1 | 2 | 2 | 1 | 2 | 2 | 3 | 0 | 3 | 2 | 2 | 3 |
| 10 | 1 | 2 | 2 | 0 | 2 | 2 | 3 | 0 | 3 | 2 | 2 | 3 |
| 11 | 1 | 2 | 2 | 1 | 2 | 2 | 3 | 0 | 3 | 2 | 2 | 3 |
| 12 | 1 | 2 | 2 | 1 | 2 | 2 | 3 | 0 | 3 | 2 | 2 | 3 |
| 13 | 1 | 2 | 2 | 1 | 2 | 2 | 3 | 0 | 3 | 2 | 2 | 3 |
| 14 | 1 | 2 | 2 | 1 | 2 | 2 | 3 | 0 | 3 | 2 | 2 | 3 |
| 15 | 1 | 2 | 2 | 1 | 2 | 2 | 3 | 0 | 3 | 2 | 2 | 3 |
| 16 | 1 | 2 | 2 | 1 | 2 | 2 | 3 | 0 | 3 | 2 | 2 | 3 |
| **Partipant** | **Sudy 2** | **CL Item 1** | **CL Item 2** | **CL Item 3** | **CL Item 4** | **CL Item 5** | **CL Item 6** | **CL Item 7** | **CL Item 8** | **Q1 Fid. Supp** | **Q2 Fid. Mon** | **Q3 Fid Rep.** |
| 1 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 3 |
| 2 | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 2 | 3 |
| 3 | 2 | 0 | 2 | 0 | 1 | 2 | 2 | 0 | 0 | 2 | 2 | 0 |
| 4 | 2 | 2 | 2 | 0 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 3 |
| 5 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 3 |
| 6 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 1 | 3 | 2 | 2 | 3 |
| 7 | 2 | 1 | 2 | 1 | 2 | 2 | 3 | 1 | 2 | 2 | 2 | 3 |
| 8 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 3 |
| 9 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 3 |
| 10 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 3 |
| 11 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 3 |
| 12 | 2 | 2 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 3 |
| 13 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 3 |
| 14 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 3 |
| 15 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 3 |
| 16 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 3 |
| **Partipant** | **Study** | **CL Item 1** | **CL Item 2** | **CL Item 3** | **CL Item 4** | **CL Item 5** | **CL Item 6** | **CL Item 7** | **CL Item 8** | **Q1 Fid. Supp** | **Q2 Fid. Mon** | **Q3 Fid Rep.** |
| 1 | 3 | 0 | 2 | 2 | 0 | 2 | 0 | 0 | 0 | 1 | 0 | 0 |
| 2 | 3 | 0 | 2 | 1 | 0 | 2 | 2 | 2 | 2 | 1 | 0 | 1 |
| 3 | 3 | 0 | 2 | 2 | 0 | 2 | 0 | 0 | 0 | 1 | 0 | 0 |
| 4 | 3 | 0 | 1 | 2 | 0 | 2 | 0 | 0 | 0 | 1 | 0 | 0 |
| 5 | 3 | 0 | 2 | 2 | 0 | 2 | 0 | 0 | 0 | 1 | 0 | 0 |
| 6 | 3 | 0 | 2 | 2 | 1 | 2 | 0 | 0 | 0 | 1 | 0 | 0 |
| 7 | 3 | 0 | 2 | 2 | 0 | 2 | 0 | 0 | 0 | 1 | 0 | 0 |
| 8 | 3 | 0 | 2 | 2 | 0 | 2 | 0 | 0 | 0 | 1 | 0 | 0 |
| 9 | 3 | 0 | 2 | 2 | 0 | 2 | 0 | 0 | 0 | 1 | 0 | 0 |
| 10 | 3 | 0 | 2 | 2 | 0 | 2 | 0 | 0 | 0 | 1 | 0 | 0 |
| 11 | 3 | 0 | 2 | 2 | 0 | 2 | 0 | 0 | 0 | 1 | 0 | 0 |
| 12 | 3 | 0 | 2 | 2 | 0 | 2 | 0 | 0 | 0 | 1 | 0 | 0 |
| 13 | 3 | 0 | 2 | 2 | 0 | 2 | 0 | 0 | 0 | 1 | 0 | 0 |
| 14 | 3 | 0 | 2 | 2 | 0 | 2 | 0 | 0 | 0 | 1 | 0 | 0 |
| 15 | 3 | 0 | 2 | 2 | 0 | 2 | 0 | 0 | 0 | 1 | 0 | 0 |
| 16 | 3 | 0 | 2 | 1 | 0 | 2 | 0 | 0 | 0 | 1 | 0 | 0 |

**Table 3.2** Cronbach alpha: CONSIDER checklist

Anova: Two-Factor Without Replication

| *SUMMARY* | *Count* | *Sum* | *Average* | *Variance* |
|---|---|---|---|---|
| Row 1 | 36 | 57 | 1.58333333 | 1.05 |
| Row 2 | 36 | 61 | 1.69444444 | 0.8468254 |
| Row 3 | 36 | 41 | 1.13888889 | 1.03730159 |
| Row 4 | 36 | 56 | 1.55555556 | 1.11111111 |
| Row 5 | 36 | 57 | 1.58333333 | 1.05 |
| Row 6 | 36 | 56 | 1.55555556 | 0.88253968 |
| Row 7 | 36 | 56 | 1.55555556 | 1.11111111 |
| Row 8 | 36 | 56 | 1.55555556 | 0.9968254 |
| Row 9 | 36 | 57 | 1.58333333 | 1.05 |
| Row 10 | 36 | 57 | 1.58333333 | 1.05 |
| Row 11 | 36 | 57 | 1.58333333 | 1.05 |
| Row 12 | 36 | 56 | 1.55555556 | 1.05396825 |
| Row 13 | 36 | 56 | 1.55555556 | 0.9968254 |
| Row 14 | 36 | 56 | 1.55555556 | 0.9968254 |
| Row 15 | 36 | 57 | 1.58333333 | 1.05 |
| | | | | |
| Column 1 | 15 | 15 | 1 | 0 |
| Column 2 | 15 | 30 | 2 | 0 |
| Column 3 | 15 | 30 | 2 | 0 |
| Column 4 | 15 | 15 | 1 | 0.14285714 |
| Column 5 | 15 | 29 | 1.93333333 | 0.06666667 |
| Column 6 | 15 | 30 | 2 | 0 |
| Column 7 | 15 | 44 | 2.93333333 | 0.06666667 |
| Column 8 | 15 | 0 | 0 | 0 |
| Column 9 | 15 | 43 | 2.86666667 | 0.26666667 |
| Column 10 | 15 | 30 | 2 | 0 |
| Column 11 | 15 | 30 | 2 | 0 |
| Column 12 | 15 | 43 | 2.86666667 | 0.26666667 |
| Column 13 | 15 | 30 | 2 | 0 |
| Column 14 | 15 | 27 | 1.8 | 0.31428571 |
| Column 15 | 15 | 30 | 2 | 0 |
| Column 16 | 15 | 13 | 0.86666667 | 0.12380952 |
| Column 17 | 15 | 28 | 1.86666667 | 0.12380952 |
| Column 18 | 15 | 29 | 1.93333333 | 0.06666667 |
| Column 19 | 15 | 30 | 2 | 0.14285714 |
| Column 20 | 15 | 15 | 1 | 0.14285714 |
| Column 21 | 15 | 38 | 2.53333333 | 0.6952381 |
| Column 22 | 15 | 30 | 2 | 0 |
| Column 23 | 15 | 30 | 2 | 0 |
| Column 24 | 15 | 42 | 2.8 | 0.6 |
| Column 25 | 15 | 45 | 3 | 0 |
| Column 26 | 15 | 0 | 0 | 0 |
| Column 27 | 15 | 29 | 1.93333333 | 0.06666667 |
| Column 28 | 15 | 29 | 1.93333333 | 0.06666667 |
| Column 29 | 15 | 1 | 0.06666667 | 0.06666667 |
| Column 30 | 15 | 30 | 2 | 0 |
| Column 31 | 15 | 2 | 0.13333333 | 0.26666667 |
| Column 32 | 15 | 2 | 0.13333333 | 0.26666667 |
| Column 33 | 15 | 2 | 0.13333333 | 0.26666667 |
| Column 34 | 15 | 15 | 1 | 0 |
| Column 35 | 15 | 0 | 0 | 0 |
| Column 36 | 15 | 0 | 0 | 0 |

ANOVA

| *rce of Variat.* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
|---|---|---|---|---|---|---|
| Rows | 7.08148148 | 14 | 0.50582011 | 5.03915663 | 7.5761E-09 | 1.71199382 |
| Columns | 487.481481 | 35 | 13.9280423 | 138.756024 | 5.952E-230 | 1.44721269 |
| Error | 49.1851852 | 490 | 0.10037793 | | | |
| | | | | | | |
| Total | 543.748148 | 539 | 0.8155409 | | | |

**Appendix IV:** Chapter 4 supplementary materials

Figure 4.1 Search strategy: Pubmed

((((("expressive language"[Title/Abstract] OR "receptive language"[Title/Abstract] OR "articulation"[Title/Abstract]) AND 2015/01/01:3000/12/31[Date - Publication] AND "humans"[MeSH Terms]) OR (("speech therapy"[Title/Abstract] OR "speech"[Title/Abstract] OR "aphasia"[Title/Abstract] OR "communication disorders"[Title/Abstract] OR "language therapy"[Title/Abstract] OR "language disorder"[Title/Abstract] OR "speech disorder"[Title/Abstract]) AND "humans"[MeSH Terms])) AND "Meta-Analysis"[Title/Abstract] AND "humans"[MeSH Terms])) OR ("2"[All Fields] AND "speech therapy"[Title/Abstract]) OR "aphasia"[Title/Abstract] OR "dysphagia"[Title/Abstract] OR "articulation"[Title/Abstract]) AND (("filter"[All Fields] OR "filter s"[All Fields] OR "filtered"[All Fields] OR "filtering"[All Fields] OR "filterings"[All Fields] OR "Filters"[All Fields]) AND ("Meta-Analysis"[Publication Type] OR "meta analysis as topic"[MeSH Terms] OR "Meta-Analysis"[All Fields]) AND ("human s"[All Fields] OR "humans"[MeSH Terms] OR "humans"[All Fields] OR "human"[All Fields]))) OR ("4"[All Fields] AND (("exercise therapy"[Title/Abstract] OR "exercise interventions"[Title/Abstract] OR "exercise treatment"[Title/Abstract]) AND "exercise"[Title/Abstract] AND "Meta-Analysis"[Title/Abstract]) AND ("Filters"[All Fields] AND "Meta-Analysis"[All Fields])) OR (("exercise therapy"[Title/Abstract] OR "exercise interventions"[Title/Abstract] OR "exercise treatment"[Title/Abstract]) AND "exercise"[Title/Abstract] AND "Meta-Analysis"[Publication Type]) OR (("physiotherapy"[Title/Abstract] OR "physical therapy"[Title/Abstract]) AND ("Meta-Analysis"[Publication Type] AND "humans"[MeSH Terms]))) AND "Meta-Analysis"[Publication Type] AND ("humans"[MeSH Terms] AND 2010/01/01:2020/06/01[Date - Publication]) AND (("rct"[All Fields] OR ("clinical trials as topic"[MeSH Terms] OR ("clinical"[All Fields] AND "trials"[All Fields] AND "topic"[All Fields]) OR "clinical trials as topic"[All Fields] OR "trial"[All Fields] OR "trial s"[All Fields] OR "trialed"[All Fields] OR "trialing"[All Fields] OR "trials"[All Fields]) OR "randomized control trial"[All Fields]) AND ("humans"[MeSH Terms] AND 2010/01/01:2020/06/01[Date - Publication])) AND ("humans"[MeSH Terms] AND 2010/01/01:2020/06/01[Date - Publication]) AND ((("mean"[All Fields] AND ("differ"[All Fields] OR "differed"[All Fields] OR "difference"[All Fields] OR "differences"[All Fields] OR "differencies"[All Fields] OR "different"[All Fields] OR "differently"[All Fields] OR "differents"[All Fields] OR "differing"[All Fields] OR "differs"[All Fields])) OR "standard mean difference"[All Fields]) AND "humans"[MeSH Terms])) AND (humans[Filter])

Table 4.1: CONSIDER Checklist evaluations of RCTs in included meta-analyses

## Meta-analysis: Avery 2012

| Item | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | Fidelity Supported? | Fidelity Monitored? | Fidelity achieved? | ROB | +/- Rx Eff? |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Balducci 2010a | 2 | 2 | 2 | 2 | 2 | 2 | 0 | 2 | 12 yes | 2 prov fid pt adh | 2 yes | mod | +/ns |
| Balducci 2010b | 2 | 1 | 0 | 0 | 2 | 0 | 0 | 0 | 5 no | 0 | 0 No | mod | +/ns |
| Cheung 2009 | 1 | 1 | 1 | 0 | 2 | 1 | -2 | 1 | 6 yes | 1 Adh | 1 rx, -2 control | mod | =/ns |
| De Greef 2010 | 2 | 2 | 2 | 1 | 2 | 1 | 0 | 1 | 10 yes | 1 adh | 1 Adh | mod | +/s |
| De Greef 2011 | 2 | 2 | 2 | 2 | 2 | 1 | 0 | 1 | 11 yes | 1 adh | 1 Adh | low | +/s |
| Gram 2010 | 2 | 2 | 2 | 1 | 2 | 1 | -2 | -2 | 10 yes | 1 adh | -2 no | low | =/ns |
| Kim 2006 | 2 | 2 | 2 | 0 | 2 | 1 | 0 | 0 | 9 yes | 1 min/adh | 0 no rep | mod | +/s |
| Ligtenberg 1995 | 2 | 2 | 1 | 1 | 2 | 1,1 | 0 | -1 | 10 yes | Yes 2 | -1 No | mod | =/ns |
| Plotnikoff 2010 | 2 | 2 | 2 | 1 | 2 | 2 | 0 | -1 | 11 yes | 2 pt fid/adh | -1 no | mod | +/ns |
| Tudor-Locke 2004 | 2 | 2 | 1 | 0 | 2 | 1 | 0 | -1 | 8 yes | 1 pt adh | -1 No | mod | +/s |

**Notes:**
Ligtenberg: Adherence at 97% for initial phase of intervention but only 62% for following phases. -1 for adherence
Plotnikoff 2011: Conference abstract, with no way of accessing a full paper to assess fidelity. Excluded.

## Meta-analysis: Bellicha 2018

| Item | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | Fidelity Supported? | Fidelity Monitored? | Fidelity achieved? | ROB | +/- Rx Eff? |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Campanha 2017 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 3 no | 0 | 0 No | mod | +/s |
| Castelo 2011 | 0 | 2 | 1 | 0 | 2 | 0 | 0 | 0 | 5 no | 0 | 0 No | mod | =/ns |
| Coen 2015 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 5 no | 2 (1 adh, 1 fid) | 1 adh | low | =/ns |
| Daniels 2017 | 0 | 2 | 1 | 0 | 2 | 0 | 0 | 1 | 5 no | 0 | 0 | mod | +/ns |
| Hassanejad 2017 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 3 no | 0 | No | low | +/s |
| Herring 2017 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 9 yes | 1 min/adh | 1 adh | low | +/s |
| Jassil 2015 | 2 | 2 | 2 | 2 | 2 | 1 | 0 | -1 | 11 yes | 1 min/adh | -1 no | high | +/s |
| Shah 2011 | 0 | 2 | 1 | 0 | 2 | 1 | 0 | -1 | 5 no | 0 | No | mod | =/ns |
| Onofre 2018 | 0 | 2 | 0 | 1 | 2 | 1 | 0 | -1 | 6 mod | 1 min/adh | No | mod | +/ns |
| Stegen 2011 | 0 | 2 | 1 | 0 | 2 | 0 | 0 | 0 | 5 no | 0 | No | mod | =/ns |

**Notes: -**Huck 2015 excluded because was a quasi-experimental study (not RCT).
-Marchesi 2014 excluded also as "a nonrandomized prospective controlled pilot trial on a selected cohort of post-bariatric patients."

## Meta-analysis: Briani 2018

| Item | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | Fidelity Supported? | Fidelity Monitored? | Fidelity achieved? | ROB | +/- Rx Eff? |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Aglamis 2008 | 1 | 2 | 0 | 0 | 2 | 0 | 0 | 0 | 5 no | 0 No | No | high | +/s |
| Broderick 2014 n | 2 | 2 | 2 | 2 | 2 | 3 | 1 | 3 | 13 yes | 3 yes | 3 yes | low | +/ns |
| Brosseau 2012 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | -1 | 6 yes | 1 min/adh | -1 no | high | -/ns |
| Buszewicz 2006 n | 1 | 0 | 0 | 0 | 1 | 1 | 0 | -1 | 3 no | 1 min/adh | -1 no | low | +/ns |
| Cadmus 2010 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | -1 | 4 no | 0 No | No | low | +/s |
| Coleman 2012 n | 2 | 2 | 1 | 2 | 2 | 2 | 0 | 2 | 11 yes | 2 yes | 2 yes | low | +/ns |
| Doi 2008 | 2 | 2 | 1 | 2 | 2 | 2 | 1 | 2 | 11 yes | 2 yes | 2 yes | mod | +/s |
| Ebenezar 2011 | 1 | 2 | 0 | 1 | 2 | 2 | 0 | 1 | 8 yes | 2 pt adh/fid | 1 adh | low | +/s |
| Hunt 2013 | 2 | 2 | 1 | 2 | 2 | 3 | 2 | 3 | 12 yes | 3 pt/pr fid/adh | 3 pt adh and fid, prov fid | Low | +/s |
| Lee 2009 | 0 | 2 | 0 | 0 | 1 | 0 | 0 | 0 | 3 no | 0 n/a | 0 n/a | low | +/ns |
| Marconcin 2018 | 2 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 4 | 0 nr | 0: adh but nr of mon. | high | +/s |
| Odole 2014 | 2 | 1 | 0 | 1 | 2 | 2 | 0 | 0 | 8 yes | 2 pt fid/adh | 0 nr | mod | +/s |
| Rejeski 2002 | 2 | 2 | 1 | 2 | 2 | 3 | -2 | 1 | 12 yes | 3 yes | 1 adh | high | +/s |
| Skou 2015 | 1 | 2 | 1 | 2 | 2 | 1 | 0 | -1 | 9 yes | 1 min/adh | -1 no | low | +/s |
| Thorstensson 2005 | 2 | 2 | 2 | 1 | 2 | 1 | 1 | 1 | 9 yes | 1 min/adh | 1 adh | low | -/ns |

**Notes:**

-Broderick: This is part 2 of a series, with part 1 giving detailed description of the intervention. Scored 1 for intervention details because can be found there but very little information in the present paper.
-Kao not eligible: quasi-experimental study
-Marconcin 2018: Describes a published protocol to aid adherence but not available and no description of the experimental intervention details or methods to assess fid/adherence described. They describe adherence as low drop out, but no monitoring.
Figure 6: Ren Lee, etc (not eligible as not comparing a phys. act or exercise based intervention. E.g.: Ren compares two Chinese medicine herbs. Only included trials comparing phys.act/exercise based interventions with QOL outcome as with fig 5.
Ebenezer: both groups receive active exercise/movement-based treatment, neither blind to treatment. Risk of bias for participants blinded rated low after consulting literature for blinding in complex intervention trials.
Ackerman-ineligible

## Meta-analysis: Cleave 2015

| Item | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | Fidelity Supported | Fidelity Monitored? | Fidelity achieved? | ROB | +/- Rx Eff? |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Fey 1993 | 2 | 2 | 1 | 1 | 2 | 2 | 0 | 2 | 10 yes | 2 pt fid/adh | 2 yes | Low | +/ns |
| Gallagher 2009 | 2 | 2 | 1 | 1 | 1 | 2 | -2 | -1 | 8 yes | 2 fid/adh | -1 no | mod | +/s |
| Robertson 1999 | 2 | 2 | 1 | 2 | 2 | 2 | 2 | 2 | 11 yes | 2 prov fid | 2 yes | high | +/ns |
| Schwartz 1985 | 1 | 1 | 0 | 0 | 2 | 0 | 0 | 0 | 3 no | 0 no report | 0 nr | high | +/ns |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Smith-Lock 2013 | 2 | 2 | 1 | 2 | 2 | 2 | 0 | 0 | 11 yes | 2 prov ad/fid | 0 no report | high | +/ns |
| Tyler 2003 | 2 | 2 | 2 | 2 | 2 | 3 | 0 | 3 | 13 yes | 3 yes | 3 yes | mod | +/s |
| Tyler 2011 | 2 | 2 | 2 | 2 | 2 | 3 | 2 | 3 | 13 yes | 3 yes | 3 yes | high | +/s |

Item Notes:

-Tyler 2011 a particularly good illustrative example of fidelity promotion, monitoring and scoring in a clinical trial.

## Meta-analysis: Depiazzi 2019

| Item | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | Fidelity Supported? | Fidelity Monitored? | Fidelity achieved? | ROB | +/- Rx Eff? |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bento 2015 | 2 | 2 | 2 | 0 | 2 | 1 | 0 | 1 | 9 yes | 1 adh | 1 adh | high | + |
| Broman 2006 | 2 | 2 | 1 | 0 | 2 | 2 | 0 | 2 | 9 yes | 2: pt fid | 2 yes | mod | + |
| Hamer 1990* | 2 | 2 | 1 | 2 | 2 | 2 | 0 | 2 | 11 yes | 2 yes fid/adh | 2 yes | high | + |
| Michaud 1995 | 2 | 2 | 2 | 0 | 2 | 2 | 0 | 0 | 10 yes | 2 yes | 0 no report | high | + |
| Mohr 2014 | 2 | 1 | 0 | 0 | 2 | 0 | 0 | 0 | 5 no | 0 no | 0 no report | mod | + |
| Moreira 2013 | 2 | 2 | 2 | 0 | 2 | 2 | 0 | 1,1 | 10 | 2: pat fid/adh | 2 yes | low | + |
| Rebold 2013 | 2 | 2 | 1 | 0 | 2 | 1 | 0 | 0 | 8 yes | 2 pt fid | 0 no report | mod | + |
| Waller 2017 | 2 | 2 | 2 | 2 | 2 | 3 | 2 | 3 | 13 yes | 3 | 3 yes | mod | + |

Item Notes:

Waller: extensive details about intervention, fidelity, etc in study protocol. Protocol cited, open access, and described in paper.

## Meta-analysis: Finch 2018

| Item | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | Fidelity Supported? | Fidelity Monitored? | Fidelity achieved? | ROB | +/- Rx Eff? |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Alhassan 2007 | 2 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 5 no | 0 no | 0 n/a | high | =/ns |
| Alhassan 2013 | 2 | 2 | 0 | 2 | 2 | 0 | 0 | 0 | 8 yes | 0 no | 0 n/a | high | =/ns |
| Alhassan 2012 | 2 | 2 | 0 | 2 | 2 | 3 | 0 | 2 | 10 yes | 2 prov fid/adh, pt adh | 2 yes | high | +/s |
| Anessi | 2 | 2 | 0 | 2 | 2 | 0 | 0 | 0 | 8 yes | 0 no | 0 n/a | mod | +/s |
| Bellows 2013 | 2 | 2 | 1 | 2 | 2 | 2 | 2 | 2 | 10 yes | 2 prov adh/fid | 2 yes | mod | =/ns |
| Bonvin 2013 | 1 | 1 | 0 | 2 | 1 | 0 | 0 | 0 | 5 no | 0 no | 0 n/a | mod | =/ns |
| Cardon 2009 | 2 | 2 | 0 | 0 | 1 | 1 | 0 | 0 | 5 no | 1 prov adh | 0 nr | mod | -/ns |
| De Bock 2013 | 2 | 2 | 2 | 2 | 2 | 2 | 0 | 2 | 12 yes | 2 prov fid/adh | 2 yes | mod | +/s |
| De Craemer | 2 | 2 | 1 | 2 | 2 | 2 | 0 | O | 11 yes | 2pt/prov fid/adh | 0 no rep | mod | =/ns |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Eliakim 2007 | 1 | 2 | 0 | 2 | 2 | 1 | 0 | 0 | 8 yes | 1 adh | 0 no rep | high | +/s |
| Finch 2014 | 2 | 2 | 1 | 2 | 2 | 2 | 0 | 0 | 11 yes | 2 prov fid/adh | 0 no | high | =/ns |
| Fitzgibbon 2010 | 2 | 2 | 2 | 2 | 2 | 2 | 0 | 1 | 12 yes | 2 prov fid/adh | 1 partial | high | +/ns |
| Jones 2011 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 8 yes | 2 prov adh/fid | 2 yes | mod | =/ns |
| O'Dwyer | 2 | 1 | 2 | 1 | 1 | 0 | 0 | 0 | 7 no | 0 no | 0 n/a | mod | =/ns |
| Pruder 2011 | 2 | 2 | 2 | 2 | 2 | 3 | 2 | 2 | 13 | 2 prov adh/fid, pt adh | 3 yes | mod | +/s |
| Reilly 2006 | 2 | 1 | 0 | 2 | 2 | 1 | 0 | 0 | 8 yes | 1 prov fid | 0 | mod | =/ns |

**Notes:**

## Meta-Analysis: Hampton 2016

| Item | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | Fidelity Supported | Fidelity Monitored ? | Fidelity achieved ? | ROB | +/- Rx Eff? |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Aldred 2004 | 1 | 1 | 1 | 0 | 2 | 0 | 0 | 0 | 5 no | 0 | 0 | low | +/ns |
| Carter 2011 | 2 | 2 | 1 | 2 | 2 | 3 | 2 | 3 | 12 yes | 3 yes pt ad/prov fid | 3 yes | mod | +/ns |
| Casenhiser 2013 | 2 | 2 | 2 | 2 | 2 | 2 | 0 | -1 | 12 yes | 2 yes prov fid/adh | -1 no | low | +/ns |
| Dawson 2010 | 2 | 2 | 0 | 2 | 2 | 1 | 0 | 0 | 9 yes | 1 prov fid | 0 nr | low | +/s |
| Drew 2002 | 2 | 2 | 1 | 1 | 2 | 0 | 0 | 0 | 8 yes | 0 no | 0 no | mod | =/ns |
| ***Goods 2013*** | 1 | 1 | 0 | 1 | 1 | 2 | 0 | 1 | 6 yes | Fid but no description how | n/a: reports prov fid not how | low | +/ns |
| Green 2010 | 2 | 2 | 1 | 2 | 2 | 3 | 2 | 2 | 12 yes | 3 yes pt ad/prov fid | 3 yes | low | +/ns |
| Hardan 2014 | 2 | 2 | 2 | 1 | 2 | 2 | 1 | 2 | 11 yes | 2 pt fid/adh | 2 yes | low | /ns |
| Kasari 2008 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 12 yes | 2 yes prov fid/adh | 2 yes | low | +/ns |
| Rogers 2012 | 2 | 2 | 1 | 2 | 2 | 3 | 2 | 3 | 12 yes | 3 yes pt /prov fid/ad | 3 yes | low | =/ns |
| Siller 2013 | 2 | 2 | 2 | 2 | 2 | 3 | 2 | 3 | 13 yes | 3 yes pt /prov fid/ad | 3 yes | low | +/ns |
| Solomon 2014 | 2 | 2 | 1 | 2 | 2 | 3 | 2 | 3 | 12 yes | 3 yes pt /prov fid/ad | 3 yes | low | +/s |
| Strain 2011 | 2 | 2 | 1 | 2 | 2 | 2 | 2 | 2 | 11 yes | 2 pt fid/adh | 2 yes | low | +/s |
| Tonge 2012 | 2 | 2 | 1 | 2 | 2 | 3 | 2 | 2 | 12 yes | 3 yes pt ad/prov fid | 3 yes | high | +/ns |
| Venker 2012 | 2 | 2 | 1 | 2 | 2 | 3 | 2 | 3 | 12 yes | 3 yes pt ad/prov fid | 3 yes | high | -/ns |
| Whalen 2010 | 2 | 2 | 2 | 1 | 2 | 0 | 0 | 0 | 9 yes | 0 no | 0 no | mod | +/ns |

Item Notes:

Tonge 2012: No formal measure of participant adherence w intervention but they describe logs and homework and providers aware of weekly adherence. Did measure participant attendance. Fid strategies and monitoring for providers.
Kasari 2008: See Kasari 2006 for intervention details/fidelity/adh assessments. 2008 reports follow up data for 2006.
Dawson 2010: Describes training and therapists reaching 85% fidelity during training/ongoing fidelity but no report of their fidelity specifically during the trial.

## Meta-Analysis: Heidlage 2020

| Item | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | Fidelity Supported? | Fidelity Monitored? | Fidelity achieved? | ROB | +/- Rx |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Aldred 2004 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 4 no | 0 no | 0 no | low | +/s |
| Boyce 2010 | 2 | 2 | 2 | 2 | 2 | 3 | 1 | 3 | 13 yes | 3 yes | 3 yes | mod | +/ns |
| Buschman 2009 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 2 no | 0 no | 0 no | low | +/s |
| Crane-Thoreson 1999 | 2 | 2 | 1 | 2 | 1 | 3 | 0 | 2 | 10 yes | 3 yes | 2 yes | high | =/ns |
| Drew 2002 | 2 | 2 | 0 | 0 | 1 | 1 | 0 | 0 | 6 no | 0 no | -1 no | high | =/ns |
| Fung 2005 | 2 | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 7 no | 0 no | 0 no | low | +/ns |
| Giralometto 1996 | 2 | 2 | 2 | 2 | 2 | 3 | 1 | 2 | 13 yes | 3 yes | 2 yes | mod | +/s |
| Guttentag 2014 | 2 | 2 | 2 | 2 | 2 | 3 | 1 | 3* | 13 yes | 3 yes | 3 yes | mod | +/ns |
| Hardan 2015 | 2 | 1 | 1 | 2 | 2 | 3 | -2 | 2 | 11 yes | 3 yes | 2 yes | mod | +/ns |
| Huebner 2000 | 2 | 2 | 1 | 2 | 1 | 3 | 2 | 3 | 11 yes | 3 yes | 3 yes | low | +/ns |
| Lonigan 1998 | 2 | 2 | 1 | 2 | 2 | 3 | -1 | -1 | 12 yes | 2 yes | -1 no | mod | +/ns |
| Pile 2010 | 2 | 2 | 0 | 2 | 2 | 3 | 0 | 3 | 11 yes | 3 yes | 3 yes | Low | =/ns |
| Roberts 2012 | 2 | 2 | 0* | 2 | 2 | 3 | 0 | 3 | 11 yes | 3 yes | 3 yes | mod | =/ns |
| Solomon 2014 | 2 | 2 | 2 | 2 | 1 | 3 | 2 | 3 | 12 yes | 3 yes | 3 yes | low | +/ns |

Item Notes:

-The primary outcomes for this MA were used: child receptive and expressive vocabulary.
-Boyce: extensive fidelity measures and scoring used
Guttentag: See supplementary material, appendix B for fidelity.
-Hardan: Intervention group achieved 92% treatment fidelity. Control group did not achieve fidelity, and the control providers' fidelity not assessed. However, considerable efforts at fidelity monitoring/enhancement intervention group, investigators, and assessors.
-Lonigan: Very high level of detail about provider training in the interventions.
-Roberts, item 5: Careful documentation of Rx/intervention group frequency but not of control group ("The amount of these services was not consistently documented by parent report.") pg. 478
Tannock: unable to access

## Meta-analysis: Hislop 2020

| Item | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | Fidelity Supported? | Fidelity Monitored? | Fidelity achieved? | ROB | +/- Rx eff? |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bennell 2014 | 2 | 2 | 2 | 2 | 2 | 1,1 | 0 | 1 | 12 yes | 2 yes | 1 adh | low | =/ns |
| Chaipinyo 2009 | 1 | 2 | 1 | 1 | 2 | 1,1 | 0 | 1,1 | 8 yes | 2 yes (1,1) | 2 yes | low | +/s |
| Olabegi 2016 | 0 | 2 | 0 | 0 | 1 | 0 | 0 | 0 | 3 no | 0 no | 0 no | mod | +/s |
| Sing 2016 | 0 | 2 | 1 | 0 | 2 | 0 | 0 | 0 | 4 no | 0 no | 0 no | mod | +/s |
| Verma 2010 | 1 | 2 | 0 | 0 | 1 | 0 | 0 | *0 | 3 no | 0 no | 0 no | mod | +/s |

Item Notes:
Ashok: Paper not available

## Meta-analysis: Howlett 2019

| Item | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | Fidelity Supported? | Fidelity Monitored? | Fidelity achieved? | ROB | +/- Rx eff? |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Aittasalo 2012 | 2 | 2 | 2 | 2 | 2 | 3 | 2 | 3 | 13 yes | 3 yes | 3 yes | mod | +/S |
| Bickmore 2013 | 2 | 2 | 1 | 1 | 2 | 1 | 1 | 1 | 9 yes | 1 min | 1 adh | mod | NS |
| Bock 2001 | 1 | 2 | 2 | 0 | 2 | 1 | 0 | 0 | 8 yes | 1 adh | 0 no rep | Mod | +/ns |
| Belanger 2013 | 1 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 4 no | 0 no | 0 no | high | NS |
| Buman 2011 | 2 | 2 | 1 | 2 | 2 | 3 | 0 | 3 | 12 yes | 3 yes | 3 yes | high | =/NS |
| Chen 1998 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 2 | 12 yes | 3 yes | 3 yes | high | =/NS |
| Dallow 2003 | 2 | 2 | 1 | 2 | 1 | 2 | 0 | 0 | 10 yes | 2: prov fid | 0* | mod | +/s |
| Hertogh 2010 | 0 | 2 | 0 | 0 | 2 | 1 | 0 | -1 | 5 no | 1 adh | -1 adh | mod | +/s |
| Kolt 2006 | 2 | 2 | 2 | 1 | 2 | 0 | 0 | 0 | 9 yes | 0 no report | 0 | mod | +/s |
| Lewis 2013 | 2 | 2 | 2 | n/a | 1 | 1 | 0 | 1 | 8 yes | 1 adh | 1 adh | high | +/s |
| Napolitano 2006 | 2 | 2 | 2 | 0 | 1 | 1 | 0 | 1 | 8 yes | 1 adh | 1 adh | mod | +/NS |
| Norton 2011 | 2 | 2 | 1 | 0 | 2 | 2 | 0 | 1 adh | 9 yes | 2: pt adh/fid | 1 adh -1 fid | mod | +/ns |
| Odenpacker 2008 | 2 | 2 | 1 | 1 | 2 | 2 | 0 | 1 adh | 10 yes | 2: pt adh/fid | 1 adh no report fid | high | +/ns |
| Rovniak 2005 | 2 | 2 | 2 | 0 | 2 | 3 | 0 | 2 | 11 yes | 3 yes | 2 pt fid/adh | high | +/ns |
| Van Hoeck 2014 | 2 | 1 | 1 | 1 | 2 | 0 | 0 | 0 | 7 yes | 0 no | 0 nr | mod | +/NS |

Notes:
Dallow: Provider's fidelity in PAR interviews monitored and checked but no reported result. Pt adherence with survey/forms low, but questionable if that is an adherence issue.
Hertogh 2010: follow-up of a sample from a previous RCT. not eligible
Nies 2006: Phys activity was not one of the interventions, but the target of educational/psycho-social Rx. Ineligible/excluded.
VanH Hoek: they report drop-out but not monitoring adherence during trial. Coaches work on motivation/barriers and facilitators but no description of monitoring adherence.

## Meta-analysis: Kunstler 2018:

| Item | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | Fidelity Supported | Fidelity Monitored ? | Fidelity achieved ? | ROB | +/- Rx Eff?/stat |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Basler 2007 | 2 | 2 | 2 | 2 | 2 | 2 | 0 | 3 | 12 yes | 2 yes | 3 yes | low | =/ns |
| De Vries 2015 | 2 | 2 | 1 | 2 | 1 | 1 | -2 | -1 | 9 yes | 1 min | -1 no | low | +/s |
| Pisters 2010 | 2 | 1 | 1 | 2 | 2 | 3 | -2 | -1 | 11 yes | 3 yes | -3 no | low | +/s |
| Van Nimwegen 2013 | 0 | 1 | 1 | 1 | 1 | 1 | -2 | -1 | 5 no | 1 min | -3 no | low | +/s |
| Wisse 2010 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | -1 | 4 no | 0 no | -1 no | high | =/ns |

Item Notes:

## Meta-analysis: Lim 2019

| Item | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | Fidelity Supported? | Fidelity Monitored? | Fidelity achieved? | ROB | +/- Rx Eff? |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bertz 2015 | 2 | 2 | 2 | 1 | 2 | 3 | 1 | 2 | 12 yes | 3 yes | 2 yes | low | +/s |
| Colleran 2012 | 2 | 2 | 2 | 1 | 2 | 3 | -2 | 1,1 | 12 yes | 3 yes | 1 partial | mod | +/s |
| Craigie 2011 | 1 | 1 | 2 | 1 | 1 | 2 | 0 | 1 | 8 yes | 2 yes | 1 adh | low | +/s |
| Dailey 2015 | 2 | 1 | 2 | 1 | 2 | 3 | 2 | 3 | 11 | 3 yes | 3 yes | mod | =/ns |
| Davenport 2011 | 2 | 2 | 2 | 0 | 1 | 2 | 0 | 1 | 9 | 2 yes | 1 adh | mod | +/s |
| deRosset 2013 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | -1 | 5 no | 1 adh | -1 no | mod | =/ns |
| Dritsa 2009 | 1 | 1 | 2 | 0 | 1 | 1 | 0 | -1 | 6 yes | 1 adh | -1 no | mod | +/s |
| Fjeldsoe 2010 | 1 | 2 | 2 | 1 | 2 | 1 | 0 | -1 | 8 yes | 1 adh | -1 no | High | +/s |
| Holmes 2018 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | -1 | 11 yes | 2 yes | -1 no | High | +/s |
| Huang 2011 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 12 yes | 1 adh | 1 adh | High | +/s |
| Keller 2014 | 2 | 2 | 1 | 1 | 2 | 1 | 0 | -1 | 9 yes | 1 adh | -1 no | mod | +/ns |
| Leermakers 1998 | 2 | 2 | 2 | 0 | 2 | 1 | 0 | -1 | 9 yes | 1 adh | -1 no | High | +/s |
| Lioret 2012 | 2 | 2 | 1 | 0 | 2 | 1 | 0 | -1 | 8 yes | 1 adh | -1 no | High | -/ns |
| Lovelady 2000 | 2 | 1 | 2 | 0 | 2 | 2 | 0 | 1 | 9 yes | 1,1 pt ad/fid | 1 adh | mod | +/s |
| Lovelady 2009 | 2 | 2 | 2 | 0 | 2 | 2 | 2 | 2 | 10 yes | 2: pt ad/fid | 2 yes | High | +/ns |
| Nicklas 2014 | 2 | 2 | 2 | 2 | 2 | 3 | 0 | 2 | 13 yes | 3 yes | 2 yes | Low | +/s |
| Ostbye 2009 | 2 | 2 | 2 | 0 | 2 | 1 | 0 | 1 | 9 yes | 1 adh | 1 adh | mod | =/ns |
| O'Toole 2003 | 2 | 2 | 2 | 1 | 2 | 2 | 0 | 1 | 11 | 2: 1 pt adh, 1 pt fid | 0 adh * | mod | +/s |
| Tripette 2014 | 2 | 1 | 0 | 0 | 1 | 0 | 0* | 0 | 5 no | 0 | 0 | High | +/s |
| Youngwanichsetha 2013 | 0 | 2 | 0 | 0 | 1 | 0 | 0 | 0 | 3 no | 0 | 0 | low | +/ns |
| Zilberman 2018 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 3 no | 0 | 0 | High | +/s |

Notes:

Berry 2015: Unable to get full text.
Bertz: secondary analysis of the LEVA trial but reports the fidelity and outcomes of the LEVA trial, so included.
Unable to identify, so excluded.
Husenovic excluded: physical activity was not an intervention in the trial, rather a recommendation.
Keller: authors describe having shown fidelity of the program in other papers but not clear where.
Kernot 2019 and Krummel 2010: Unable to get full texts.
Khodabandeh 2017: No citation in the paper, only data. Excluded.
Lovelady 1995: Data reported but no citation in paper. Excluded.
Maturi 2011: " " " " "
McCrory 1999 " " " "
***McIntyre***: small pilot study with greater focus on feasibility, Excluded.
O'Toole: adherence achieved at first follow-up for experimental group (82%), but 62% for control
Parsa 2017: Doesn't include PA as a primary intervention. excluded.
Tripette: Retroactively looked for adherence using data from game consoles, but not during intervention, so 0 for fid/adh monitoring
Wiltheiss 2013: Doesn't include PA as a primary intervention. excluded.
Zourdalani 2015: Data reported but no citation in paper. Excluded.
Krummel 2010: Unable to access after several attempts. Excluded.

## Meta-analysis: McMichan 2018

| Item | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | Fidelity Supported? | Fidelity Monitored? | Fidelity achieved? | ROB | +/- Rx Eff? |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Cui 2018 | 2 | 1 | 2 | 1 | 2 | 2 | 0 | 2 | 10 yes | 2 yes | 2 yes | mod | =/ns |
| Whittemore 2013 | 1 | 1 | 0 | 2 | 2 | 1,1 | -2 | 3 | 8 yes | 2 yes | 3 yes | mod | +/s |
| Ghaffari 2012 | 1 | 1 | 1 | 0 | 1 | 0 | -2 | 0 | 4 no | -2 no | no | high | +/ns |
| Schwarzer 2010 | 1 | 1 | 2 | 0 | 1 | 0 | 0 | 0 | 5 no | 0 no | no | high | +/ns |
| Sprujit-Metz 2008 | 2 | 1 | 0 | 0 | 2 | 0 | 0 | 0 | 5 no | 0 no | no | low | +/s |

**Item Notes:** Adapted version of the EPHPP tool for ROB

## Meta-analysis: Meta-Analysis: Nye

| Item | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | Fidelity Supported | Fidelity Monitored ? | Fidelity achieved ? | ROB | +/- Rx Eff? |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Craig 1996 | 2 | 2 | 2 | 2 | 2 | 2 | 0 | 2 | 12 yes | 2 adh/fid | 2 yes | high | +/S |
| Franken 2005 | 2 | 2 | 2 | 2 | 2 | 2 | 0 | 2 | 12 yes | 2 pt fid/adh | 2 yes | mod | +/ns |
| Harris 2002 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 3 no | 0 no report | 0 no rep | low | -/ns |
| Harrison 2004 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 3 no | 0 no rep | 0 no | mod | -/ns |
| Jones 2005 | 2 | 2 | 1 | 1 | 2 | 2 | -1 | 2 | 10 yes | 2 pt fid/adh | 2 yes | low | +/s |
| Lattermann 2008 | 1 | 1 | 1 | 2 | 5 | *1 | 0 | 0 | 9 yes | 0: adh rep but no mon. | 0 no | mod | +/s |
| Riley 2000 | 1 | 1 | 0 | 1 | 2 | 0 | 0 | 0 | 5 no | 0 no rep | 0 no | mod | +/ns |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ryan 1995 | 2 | 2 | 1 | 2 | 2 | 2 | 1 | 0 | 11 yes | 2 prov ad/fid | 2 yes* | high | +/ns |

**Notes:** Ryan 1995: 77% fidelity achieved *by end of year.*

## Meta-analysis: Scott 2020

| Item | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | Fidelity Supported? | Fidelity Monitored? | Fidelity achieved? | ROB | +/- Rx Eff? |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Adams 2017 | 2 | 2 | 2 | 0 | 2 | 2 | 0 | 2 | 10 | 2 adh/fid | 2 | mod | +/s |
| Al Majid 2015 | 1 | 2 | 2 | 1 | 1 | 2 | 0 | 2 | 8 yes | 3 pt/prov | 2 | mod | =/ns |
| | | | | | | | | | | | | | |
| Corneya 2008 | 2 | 2 | 2 | 1 | 2 | 2 | 0 | 1 | 11 yes | 1 adh | 1 adh | low | +/s |
| Corneya 2009 | 2 | 2 | 2 | 1 | 2 | 2 | 0 | 1 | 10 yes | 1 adh | 1 adh | High | +/s |
| De Luca 2016 | 2 | 2 | 2 | 1 | 2 | 0 | 0 | 0 | 9 yes | 0 no | 0 no | mod | +/s |
| Do 2015 | 2 | 2 | 1 | 1 | 2 | 1,2 | -2 | -1 | 9 | 2: 1 pt adh, 1 pt fid | 0 adh * | High | +/s |
| Dronkers 2010 | 2 | 2 | 1 | 1 | 2 | 1 | 0 | 1 | 9 | 1 adh | 1 adh | low | |
| Edvarsen 2015 | 2 | 2 | 2 | 2 | 2 | 2 | 0 | 2 | 12 | 2 fid | 2 fid | low | +/s |
| Giallauria 2016 | 1 | 1 | 1 | 0 | 2 | 0 * | 0 | 0* | 5 no | 0 no report | 0 * | mod | +/s |
| Hornsby 2014 | 2 | 2 | 2 | 1 | 2 | 2 | 0 | 2 | 11 yes | 2 pt fid/adh | 2 yes | low | +/s |
| Hvid 2016 | 1 | 2 | 1 | 0 | 2 | 1, 1 | 0 | 2 | 8 yes | 1,1 pt ad/fid | 3 (pt adh/fid | High | +/s |
| Jones 2014 (safety) | 1 | 1 | 0 | 0 | 2 | 1 | 0 | 1 | 6 yes | 1 adh | 1 adh | mod | +/s |
| Mehnert 2011 | 2 | 2 | 1 | 1 | 2 | 0 | 0 | 0 | 8 yes | 0 | 0 | High | +/ns |
| Nuri 2012 | 1 | 2 | 1 | 0 | 2 | 0 | 0 | 0 | 6 | 2 | 0 | mod | +/s |
| Nuri 2016 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 4 no | 0 no | 0 no | mod | +/s |
| Pinto 2013 | 2 | 2 | 2 | 2 | 2 | 3 | 0 | 3 | 13 | 3 pt/prov | 3 | low | +/s |
| Scott 2013 | 2 | 2 | 2 | 1 | 2 | 1 | 0 | 1 | 10 | 1 adh | 1 adh | low | +/s |
| Segal 2001 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 4 | 1 adh | 0 no | low | +/s |
| Stefanelli 2013 | 2 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 5 | 0 | 0 | mod | +/s |
| Swisher 2015 | 2 | 2 | 2 | 2 | 2 | 1 | 0 | 1 | 11 | 1 adh | 1 adh | low | +/s |
| Thorstensson 2005 | 1 | 0 | 2 | 1 | 1 | 1 | -1 | 0 | 5 | 1 adh | -1 no | High | +/s |

Notes: Giallauria: No report of adh/fid measurement but report that participants' average exercise intensity met the target they were aiming for (VO2Max). However, as was an average, unable to tell range, and so 0 for fidelity. Excluded: Jones 2013-unable to get full report. Broderick already included in Briani Meta-analysis, Casala, not eligible. Unable to access Banergee and Cormie.

## Meta-analysis: Salamh 2019

| Item | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | Fidelity Supported? | Fidelity Monitored? | Fidelity achieved? | ROB | +/- Rx Eff? |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bailey 2017 | 2 | 2 | 1 | 0 | 2 | 2 | 0 | 2 | 9 yes | 2 Prov. fid | 0 nr | low | +/s |
| Laudner 2014 | 2 | 2 | 2 | 0 | 2 | 1 | 2 | 2 | 9 yes | 1 prov. fid | 2 yes | mod | +/s |
| Menske 2010 | 1 | 2 | 0 | 1 | 2 | 1 | 0 | 1 | 7 yes | 1 adh | 1 adh | high | =/ns |
| Moore 2011 | 2 | 2 | 0 | 0 | 2 | 0 | 0 | 0 | 6 yes | 0 no | 0 no | mod | +/s |
| Salamh 2015 | 2 | 2 | 0 | 1 | 2 | 0 | 0 | 0 | 7 yes | 0 no | 0 n/r | high | +/s |
| Yang 2014 | 2 | 2 | 1 | 0 | 2 | 0 | 0 | -1 | 7 yes | 0 no | 0 n/a | high | +/s |

Notes:

Laudner 2014: extensive training to ensure assessor fidelity. Methods to ensure Rx fid by delivering standard treatment, but not truly fidelity assessment. Scored 2 but can down score to 1/partial

Sauers 2007: Uncontrolled study. Ineligible.

## Meta-analysis: Salamh 2016

| Item | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | Fidelity Supported? | Fidelity Monitored? | Fidelity achieved? | ROB | +/- Rx Eff? |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Crossley 2016 | 2 | 2 | 2 | 2 | 2 | 3 | 1 | 1 | 12 yes | 2 pt adh/fid | 2 yes | low | =/ns |
| Deyle 2000 | 2 | 1 | 2 | 0 | 2 | 2 | 0 | 2 | 9 yes | 2 pt adh/fid | 2 yes | low | +/s |
| Dwyer 2015 | 2 | 2 | 2 | 2 | 2 | 2 | 0 | 1 | 12 yes | 2 pt adh/fid | 1 adh | low | =/ns |
| Deyle 2005 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 2 | 12 yes | 2 pt adh/fid | 2 yes | low | +/s |
| Pollard 2008 | 0 | 2 | 1 | 0 | 2 | 0 | 0 | 0 | 5 no | 0 | 0 n/a | high | +/s |

Notes:

Kappetijin 2014: A parallel group observational study: not eligible/excluded.

## Meta-analysis: Young 2018

| Item | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | Fidelity Supported ? | Fidelity Monitored ? | Fidelity achieved | ROB | +/- Rx Eff? |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Carmack 1995 | 2 | 2 | 1 | 0 | 2 | 1 | 0 | 1 | 8 yes | 1pt adh | 1 adh | mod | +/s |
| Groussard 2015 | 2 | 2 | 2 | 1 | 2 | 1 | 0 | 1 | 10 yes | 1 pt fid? | 0 nr | mod | +/ns |
| Painter 2002 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 2 | 11 yes | 2 pt adh/fid | 2 adh/?fid | low | +/s |
| Reboredo 2010 | 2 | 2 | 2 | 0 | 2 | 1 | -1 | -1 | 10 yes | 2 pt adh/?fid | -1 no, fid/adh | mod | +/ns |
| Koh 2010 | 2 | 2 | 2 | 0 | 2 | 2 | -1 | -1 | 5 no | 2 pt adh/?fid | -1 no, fid/adh | high | =/ns |

**Notes:**
Painter, Reboredo, Groussard: HR monitoring to monitor exercise intensity, could be seen as fidelity to intended level of exercise, but debatable if this was intention or is sufficient?

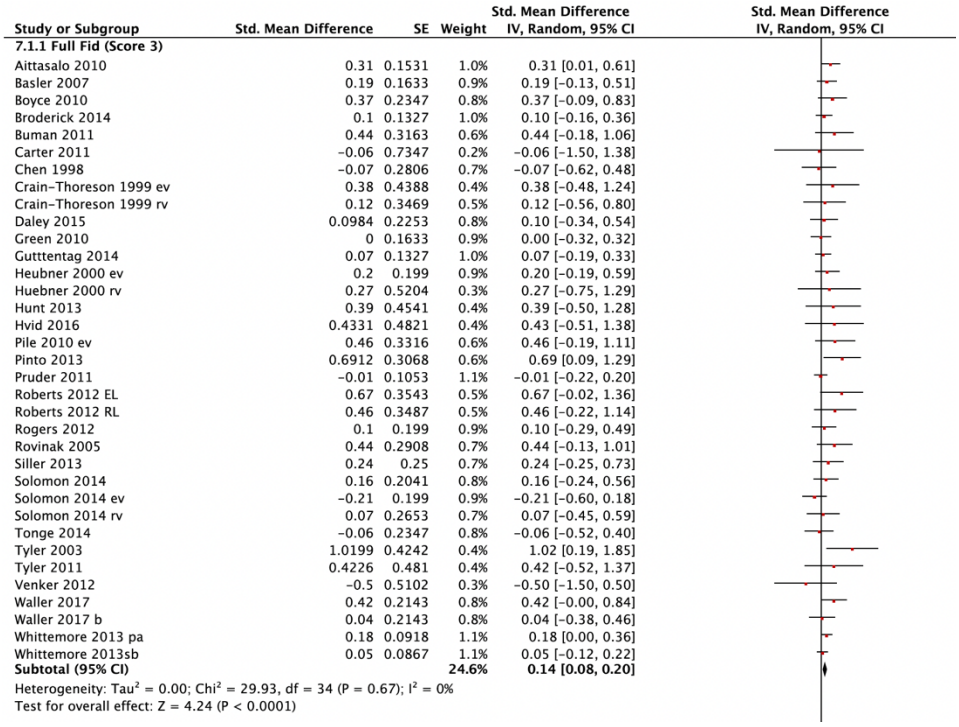
## Meta-analysis: Ward 2019

| Item | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | Fidelity Supported? | Fidelity Monitored? | Fidelity achieved? | ROB | +/- Rx eff? |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bauman 2012 | 2 | 2 | 2 | 2 | 2 | 1 | 0 | -1 | 11 yes | 1 adh | -1 no | mod | +/ns |
| Borghi-S 2015 | 2 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 5 no | 0 | 0 | low | +/s |
| Borghi-S 2009 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 4 no | 0 nr | 0 | mod | +/s |
| Duruturk 2015 | 2 | 2 | 2 | 1 | 2 | 1 | 0 | 1 | 10 yes | 1 adh | 1 adh | low | +/ns |
| Emery 1998 | 2 | 2 | 1 | 1 | 2 | 1 | 0 | 1 | 9 yes | 1 adh | 1 adh | low | +/s |
| Larson 1999 | 2 | 2 | 2 | 1 | 2 | 2 | 0 | 2 | 11 yes | 2 | 2 | high | +/s |
| Lake 1990 | 2 | 2 | 1 | 1 | 2 | 1 | 0 | 1 | 9 yes | 1 adh | 1 adh | mod | +/s |
| Reardon | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 4 no | 0 | 0 | low | +/ns |
| Ries 1995 | 2 | 1 | 2 | 1 | 2 | 0 | 0 | 0 | 8 yes | 0 nr | 0 | low | +/s |
| Troosters 2000 | 1 | 2 | 1 | 1 | 1 | 1 | 0 | 1 | Yes 7 | 1 adh | 0 | low | +/s |
| Wijkstra 1996 | 2 | 2 | 1 | 2 | 2 | 2 | 0 | 1 | Yes 11 | 1 adh | 0 n/r | mod | +/s |
| Zambom-F 2015 | 2 | 2 | 2 | 0 | 2 | 1,1 | 0 | 1 | 8 yes | 1 pt adh, 1 pt fid | 1 adh | low | +/s |

**Notes:** Unable to access: Gohl 2006 (in German, unable to access or pay for access).

Figure 4.2 Forest plot, trials with full fidelity vs trials not reporting fidelity

*Note: this forest plot has been broken into two pages due to its size




a. RCTs with fidelity and adherence (full fidelity)

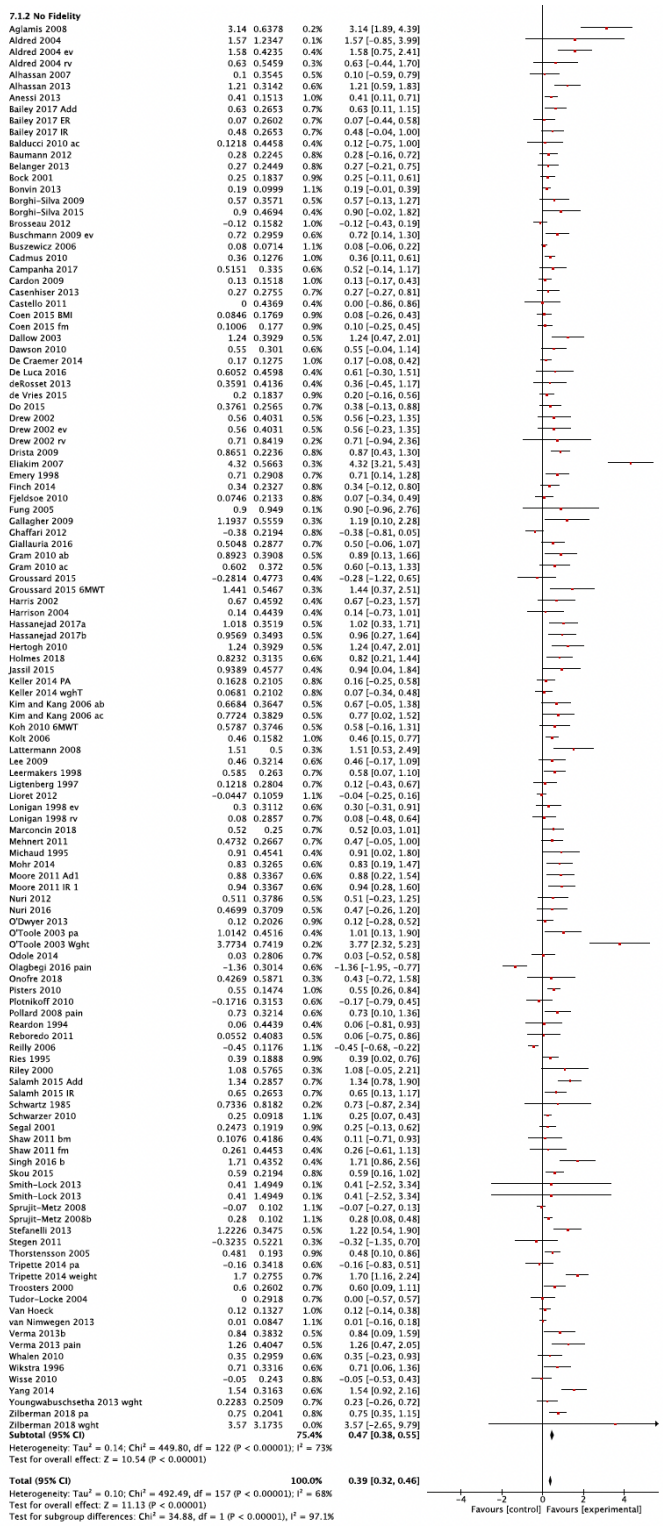



**Table 4.2 continued:** Forest plot, trials with full fidelity vs trials not reporting fidelity

**b.** RCTs not reporting fidelity or fidelity monitoring, or reporting fidelity not maintained.

**Appendix V:** Chapter 5-ACL-SNNAP supplementary materials

Figure 5.1 Surgical case-report form[349]

Study Number A C L _ _

## ACL SNNAP Operation Form

Instructions for completion:

- Completed by ACL SNNAP researcher (Surgeon/Research Nurse/Assistant) for patients undergoing an ACL reconstruction operation.
- Surgical data **must** be collected by the surgeon (or via liaison with surgeon) at the time of operation.
- Once completed please submit data to ACL SNNAP database.

| Study knee | Date of operation: | Surgeon: | Hospital: | Day Case? |
|---|---|---|---|---|
| Right ☐ Left ☐ | DD/MON/YYYY | | | Yes ☐ No ☐ |

### Section 1 - Time

| Operation time (knife to skin – Dressings on) | Theatre time (into Anaesthetic Room – out of theatre) |
|---|---|
| minutes | minutes |

### Section 2 - Staffing *(Please state the number of each present during the operation)*

| | | | | | |
|---|---|---|---|---|---|
| **Anaesthetic** | Consultant ☐ | Fellow ☐ | Registrar ☐ | SHO ☐ | Student ☐ |
| **Surgical** | Consultant ☐ | Fellow ☐ | Registrar ☐ | SHO ☐ | Student ☐ |
| **Nursing** | Sister ☐ | Staff Grade ☐ | OPD ☐ | HCA ☐ | Student ☐ |

### Section 3 - Anaesthetic *(Tick all that are applicable)*

| | | | | |
|---|---|---|---|---|
| General anaesthetic ☐ | Periarticular LA infiltration ☐ | Epidural ☐ | Sciatic Block ☐ | Femoral Block ☐ |
| Spinal ☐ | Other *please specify* | Please state ASA Grade ☐ | | |

### Section 4 - Incision and approach

| | |
|---|---|
| Arthroscopic ☐ | Open ☐ |

### Section 5 - Articular cartilage

Is the articular cartilage **normal** throughout? Yes ☐ No ☐
If No, please state reason below:

| | |
|---|---|
| Patella damage size cm² ☐ (ICRS) Grade ☐ | Trochlea damage size cm² ☐ (ICRS) Grade ☐ |
| Medial femoral damage size cm² ☐ (ICRS) Grade ☐ | Lateral femoral damage size cm² ☐ (ICRS) Grade ☐ |
| Medial plateau damage size cm² ☐ (ICRS) Grade ☐ | Lateral plateau damage size cm² ☐ (ICRS) Grade ☐ |

### Section 6 - Procedure: ACL reconstruction

Please document the type of graft used.

| | | |
|---|---|---|
| Hamstring tendon ☐ | Patella tendon ☐ | Other *please state* |

Was a Notchplasty Performed? Yes ☐ No ☐

| | | | | |
|---|---|---|---|---|
| Please check the appropriate **Femoral Tunnel Drilling Technique**: | Outside-in ☐ | Trans-tibial ☐ | AM Portal ☐ | All inside ☐ |
| Please check the appropriate **Tibial Tunnel Drilling Technique**: | | Outside-in ☐ | Inside-out ☐ | All inside ☐ |

ACL SNNAP_OperationForm_V1.0_23Aug2016

## Section 7 - Additional surgery

Was additional surgery performed? Yes ☐ No ☐
If Yes, please indicate below

| | | |
|---|---|---|
| Partial Medial Meniscectomy ☐ | Medial Meniscal Repair ☐ | Medial Replacement (Synthetic) ☐ |
| Partial Lateral Meniscectomy ☐ | Lateral Meniscal Repair ☐ | Lateral Replacement (Synthetic ☐ |
| Medial Transplant (Allograft) ☐ | Saucerisation Medial Discoid Meniscus ☐ | Posterolateral Corner Surgery ☐ |
| Lateral Transplant (Allograft) ☐ | Saucerisation Lateral Discoid Meniscus ☐ | Collateral Ligament Surgery ☐ |
| Extensor Mechanism Surgery ☐ | Articular Cartilage Surgery ☐ | PCL Surgery ☐ |
| Other ☐ *please specify* | | |

## Section 8 - Medial & lateral meniscal status at end of procedure

| | | | |
|---|---|---|---|
| Normal **Medial** Meniscus ☐ | 2/3 Remaining ☐ | 1/3 Remaining ☐ | <10% Remaining ☐ |
| Normal **Lateral** Meniscus ☐ | 2/3 Remaining ☐ | 1/3 Remaining ☐ | <10% Remaining ☐ |

## Section 9 - Intra- operative complications

Were there any intra-operative complications experienced? Yes ☐ No ☐
If there were **intra-operative complications**, please indicate below:

| | | | |
|---|---|---|---|
| Patella fracture ☐ | Patella tendon avulsion ☐ | Femoral tunnel blowout ☐ | Haemorrhage ☐ |
| Vascular Injury ☐ | Nerve damage ☐ | Ligament injury ☐ | Contralateral graft harvest ☐ |
| Graft failure ☐ | Death ☐ | Other *please specify* | |

## Section 10 - Post-operative complications

Were there any post-operative complications experienced? Yes ☐ No ☐
(from time of operation until discharge)
If there were **post-operative complications**, please indicate below:

| | | |
|---|---|---|
| DVT confirmed by<br>Blood test ☐<br>Radiology ☐ | Would infection confirmed by Microbiology ☐ | PE confirmed by<br>Blood test ☐<br>Radiology ☐ |
| Blood loss requiring a blood transfusion? ☐<br><br>If yes, how many units given? ☐ | Death ☐ | Please add any additional information about any complications: |

## Section 11 - Length of hospital stay

Patient's admission date D D / M O N / Y Y Y Y

Patient's discharge date D D / M O N / Y Y Y Y

Was the patient admitted to ICU and/or HDU during their stay in hospital? Yes ☐ No ☐

If yes, please specify the time spent in ICU and/or HDU

ICU ☐☐☐ hours HDU ☐☐☐ hours

Figure 5.2: Rehabilitation (non-surgical management arm) case report form[349]

Study Number | | | | | |

# ACL SNNAP Rehabilitation Treatment Form

Instructions for completion:

- To be completed for all participants.
- Section 1 includes 6 treatment logs to document detail on treatment sessions attended (corresponding to the minimum requirement for the study). However, please record details on all treatment sessions attended, additional treatment logs are included in the appendix.
- Please complete section 2 when the participant is discharged from physiotherapy.
- When completed, please enter data onto the ACL SNNAP database.

## Section 1 – Treatment sessions

Details to be completed for every session attended.

Session No: | | |

Date: D D / M M / Y Y Y Y

Duration of session: | |

Name of physiotherapist:

Staff grade: (please tick one box only) 5 ☐ 6 ☐ 7 ☐ 8 ☐ Other (please state) ☐

**Type of session:**

Individual one to one treatment ☐ Group based, such as knee class ☐

**Content of the session** (please tick all that apply):

| | | | |
|---|---|---|---|
| Advice and education | ☐ | Home exercises (instruction/review) | ☐ |
| Supervised exercises (strengthening) | ☐ | Gait re-education | ☐ |
| Supervised exercises (stretching) | ☐ | Supervised exercises (proprioception) | ☐ |
| Supervised exercises (sport specific) | ☐ | Hydrotherapy | ☐ |
| Other (If other please state below) | ☐ | | |