Optimizing management of musculoskeletal pain disorders in primary physiotherapy care (SupportPrim)

Statistical Analysis Plan

Version [Number: 2.0]

Administrative information

Sponsor name	Norwegian University of Science and Technology (NTNU)
Sponsor address	Postboks 8905, MTFS, 7491, Trondheim, Norway
EudraCT number / REC no	Regional Committee for Medical Research Ethics - Mid Norway ref. 49308
Trial title	Optimizing management of musculoskeletal pain disorders in primary physiotherapy care
Trial ID	
Trial registration number	ISRCTN17927832

SAP and protocol version

SAP version / date	1.0 / 14.06.2022
Protocol version	1.0. Clinical decision support system for personalized care in patients with musculoskeletal disorders – pilot and RCT study in primary care physiotherapy (Approved 08/09/2020, Regional Committee for Medical Research Ethics- Mid Norway)

SAP revision history

Protocol version	SAP version	Date changed	Description and reason for change
1.0	2.0	28.02.23	For clarification we have defined how a 30 % increase is calculated and given an example of the calculation. Added a sensitivity analysis for the second primary outcome (PSFS), using minimum 3 points improvement from baseline as definition of clinically important improvement at 3 months. Added that the sample size based on the GPE primary outcome of 15 % between group difference in proportion with improvement also holds for the other primary outcome (PSFS). Removed the redundant alternative power calculation due to this change. In addition, we have corrected some minor details of no relevance for the analysis plan.

Roles and responsibilities

Main author:	Fredrik Granviken, Department of Public Health and Nursing, Norwegian University of Science and Technology (NTNU).
Statistician:	Melanie Rae Simpson, Department of Public Health and Nursing, Norwegian University of Science and Technology (NTNU).
Project manager:	Ottar Vasseljen, Department of Public Health and Nursing, Norwegian University of Science and Technology (NTNU).
Main supervisor:	Ingebrigt Meisingset, Department of Public Health and Nursing, Norwegian University of Science and Technology (NTNU).
Project management AI:	Kerstin Bach, Department of Computer Science, Norwegian University of Science and Technology (NTNU).
Contributor to SAP:	Danielle van der Windt, School of Medicine, Keele University.

Signatures:

_____	_____	_____
Fredrik Granviken	**Melanie Rae Simpson**	**Ottar Vasseljen**
Main author	Statistician	Project Manager

Abbreviations

CDSS	Clinical decision support system
CONSORT	Consolidated Standards of Reporting Trials
EQ-5D	EuroQoL - health-related quality of life
FYSIOPRIM	Physiotherapy in Primary Health Care
GPE	Global perceived effect
HSCL	Hopkins Symptom Checklist
ICC	Intra-class correlation coefficient
ICH	The International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use
MAR	Missing at random
MSD	Musculoskeletal disorders
MSK	Musculoskeletal
MSK-HQ	The Musculoskeletal Health Questionnaire
NTNU	Norwegian University of Science and Technology
PSEQ	Pain Self Efficacy
PSFS	Patient Specific Functional Scale
RCT	Randomized controlled trial
SAP	Statistical Analysis Plan
SD	Standard deviation

Table of contents

1. Introduction

This Statistical Analysis Plan (SAP) should be read in conjunction with the protocol version 1.0 approved 08/09/2020 by the regional Committee for Medical Research Ethics- Mid Norway, "Clinical decision support system for personalized care in patients with musculoskeletal disorders – pilot and RCT study in primary care physiotherapy". The information available here provides a more detailed description of the "Statistical analysis" section.

The structure of this SAP follows the guidelines provided by Gamble et al (2017) (Available at: https://jamanetwork.com/journals/jama/fullarticle/2666509) and the checklist available from: http://lctc.org.uk/SAP-Statement

All analyses will be reported according to CONSORT extension for Cluster Trials updated 2012 (Campbell et al, 2010) and ICH E9 guidelines on Statistical Principles in Clinical Trials.

1.1 Purpose and scope of the statistical analysis plan

This document details the proposed analysis of the main paper(s) reporting results from the trial: *Optimizing management of musculoskeletal pain disorders in primary physiotherapy care (SupportPrim)*. Any deviations from the analyses outlined in this SAP will be described and justified in the final report of the trial, including the inclusion of any analyses suggested by journal editors and referees. Modifications will be carefully considered and, as far as possible, will follow the broad principles set out here.

First and foremost, this SAP describes the analysis of the primary and secondary outcomes. Subsequent exploratory analyses are also expected to follow the broad principles of this SAP but are not described in detail here.

The details presented here shall not prohibit accepted practices, such as data transformation prior to analysis. When possible, such data management and modelling decisions will be undertaken prior to revealing the treatment allocation.

The final analysis strategy will be available on request when the principal papers are submitted.

1.2 Background and rationale

Musculoskeletal disorders (MSDs) are the number one cause of years lived with disability and reduced health worldwide. In Norway, every fourth patient in primary care suffers from MSDs. Treatment effects are however modest, and knowledge of best practice limited. The SupportPrim project will address these challenges. To optimize person-centered care, we will employ methods from artificial intelligence in terms of Case-Based Reasoning to build a clinical decision support system (CDSS) based on patient data already collected in primary care physiotherapy. Case-Based Reasoning aims to solve new problems based on solutions to similar problems in the past. In other words, previous MSD cases will be used to help similar cases in the future, just as humans learn from their own experience. We will then assess the effectiveness of the CDSS in physiotherapy practice in a cluster-randomized controlled trial.

1.3 Objectives

1.3.1 Primary objective

To optimize person-centered care, we will build a CDSS for primary care physiotherapy. The main objective in this cluster randomized trial is to evaluate the effectiveness of the decision support system in physiotherapy practice. This will be evaluated with two primary outcomes: 1. Patients' assessment of their condition measured by Global perceived effect (GPE). 2. Assess if a clinically important improvement in function (PSFS) is achieved throughout the first three months.

1.3.2 Secondary and exploratory objectives

We will also evaluate the effect of the use of the CDSS on measures of pain and quality of life. A detailed list of secondary outcomes is listed in §5.2.1.

2. Study design

2.1 Study design

The trial is a cluster randomised multicentre clinical trial investigating the effectiveness of a decision support system in primary care physiotherapy practice in Norway. Treatment allocation is based on simple randomisation of physiotherapists with a 1:1 ratio to either CDSS added on to usual care or usual care only. The analyses are blinded. We do not anticipate contamination between therapists. Mostly, there are only one physiotherapist included in the study from any given practice, therapists are educated not to discuss the use of the CDSS with other therapists in the study and patients are only treated by one physiotherapist in the study period.

2.2 Randomization and treatment assignment

A web-based trial management system, administered by the Clinical Research Unit Central Norway, Faculty of Medicine and Health Sciences of the Norwegian University of Science and Technology, was used in randomization. Forty-four (44) physiotherapists were randomized in 1:1 ratio to the control and intervention group, ensuring an even number in each group. The randomization was stratified on type of physiotherapist defined by whether they were a general physiotherapist or manual therapist (which requires additional educational training in Norway). After randomization one of the therapists withdrew from the trial due to illness and was replaced with another therapist from a randomized waiting list also stratified on type of physiotherapist.

2.3 Determination of sample size

Sample size calculations based on the clustersampsi command in Stata for cluster-randomized, controlled trials. We had sufficient data to make informed estimates about the GPE outcome, and the sample size calculations have been based first and foremost on this outcome. The calculation showed that 280 patients and 20 clusters (physiotherapists) are necessary in each arm to detect a difference of 15% in proportions of patients who "improved" at 3-months follow-up (55% in control versus 70% in the CDSS/experimental group). These calculations are based on a power of 80%, alpha level of 0.05, intra-class correlation coefficient (ICC) of 0.05, and an average cluster size of 14 patients per therapist. Suggestive values for ICC were obtained from available FYSIOPRIM data (Evensen et al., 2018), and the same data suggest that the proportion of patients reporting

"improved" after usual care physiotherapy was 50-58% in the target groups (unpublished data). To account for a 20-25 % drop-out, 18 patients were to be included in each cluster.

The sample size would also be sufficient to detect a 15% between group difference in proportion with clinically important improvement for the other primary outcome, i.e., participants who experience a clinically important improvement in their functional status on the PSFS scale (30%).

These initial sample size calculations were based on the comparison between treatment arms at a single time-point and did not take into account potential variability in cluster-size. The planned analysis strategy will include repeated measures for both primary outcomes using mixed logistic regression models and the above sample size calculation can therefore be considered conservative. We have also confirmed that the coefficient of variation in cluster size can be as high as 0.4 without any loss of power using the above calculation strategy.

2.4 Framework

The study is designed as a RCT to test superiority assessing the effectiveness of the decision support system in addition to usual care compared to usual care only.

2.5 Statistical interim analyses and stopping guidance

No interim analysis is planned for the trial.

2.6 Timing of final analysis

The primary outcome will be assessed by 3 months follow-up, and the main secondary outcomes will be assessed at timepoints up to 3-months. As such, all data from baseline, 2 weeks, 4 weeks, 8 weeks, and 3 months will be exported, analysed, and published together. Follow-up data are available as of March 2022.

Analyses of treatment effect at 6 months- and 12 months follow up will be analysed and published later. These analyses will follow the overall principles outlined in this SAP.

2.6.1 Timing of endpoint assessments

The 3 months follow up data were collected by sending out the questionnaire to the patients 3 months after baseline consultation. If not answered, we sent 3 reminders during the following two weeks. In addition, any participants who had still not answered, were contacted per telephone.

3. Statistical principles

3.1 Confidence intervals and p-values

Estimates will be presented as mean differences or odds ratios, and their precision with 95% CI without p-values, as recommended in the literature (Wasserstein & Lazar, 2016; Wasserstein et al, 2019). Where relevant, e.g., interactions in subgroup analyses, p-values will be reported numerically.

3.2 Uptake, protocol deviations and protocol violations

The patients were asked if they could be invited to the study by their therapists. They got information about the study by mail, consented to participate and answered the baseline questionnaire before the first consultation. At the first consultation the physiotherapist made the final decision about enrolling the patient to the study considering inclusion and exclusion criteria.

3.2.1 Uptake

In the intervention arm were given access to a CDSS for shared decision making and optimal management of patients in addition to usual care. We define uptake to the intervention as use of the CDSS. We use Matomo analytics to see how much the system is used by the individual clinicians. We will define high use and low use based on number of clicks in the dashboard in the CDSS. Number of clicks will be presented. This will provide basis for per protocol analysis.

3.2.2 Protocol deviations

Technical failures with server-downtime that causes trouble for clinicians to use the system will be registered. The time window for answering the primary endpoints at 3 months will be from 3-5 months and we will report the variation in follow-up time for this endpoint.

3.3 Analysis populations

3.3.1 Intention to Treat (or Full Analysis Set)

The intention-to-treat principle will be used for analysing effects of the study intervention. We will include all randomised therapists allocated to either intervention or control group and their included participating patients.

3.3.2 Per-Protocol Analysis Set

A per-protocol analysis will be conducted for the primary outcomes, excluding clinicians (clusters) with low number of clicks (to be decided, e.g., less than 10% of the upper tenth percentile) when using the CDSS and participants who completed the 3-month follow-up outside of the identified time window (3-5 months). We will consider a separate analysis excluding only the participants who were followed-up beyond 5 months. These analyses will be completed after analyses based on the intention-to-treat analysis are finished, as there is a risk that the per-protocol analyses will not be blinded.

3.3.3 Subgroup definitions

For exploratory analysis we hypothesize that physiotherapists might use the CDSS in a better way when they have used the system for a while and those having higher uptake with the CDSS might have larger effect with patients. Complex patients, with more pain sites, higher severe symptoms and/or poorer prognosis might have larger effect of the CDSS. Therefore, we will do exploratory analysis of these pre-specified subgroups:

- The first 9 recruited patients compared to the patients numbered 10 and above for each physiotherapist.
- Patients with two or less musculoskeletal (MSK) pain sites versus patients with 3 or more MSK pain sites at baseline.
- Patients with less than 1.85 points versus patients with 1.85 points or higher in Hopkins Symptom Checklist (HSCL-10) at baseline.
- Patients with scores <8 vs ≥8 on the work ability scale at baseline
- Based on Matomo data (clicks in the CDSS); physiotherapists with number of clicks in the lower vs the upper quartile in the intervention group to see if higher uptake of the CDSS is important

- Comparing five phenotype groups as defined by our previous work (Meisingset et al., 2020).

3.4 Allocation concealment, blinding and order of analysis

3.4.1 Allocation concealment

Treatment allocation is based on simple randomization among clusters (physiotherapists) with a 1:1 ratio to CDSS in addition to usual care or usual care only. This was administered by the Clinical Research Unit Central Norway, Faculty of Medicine and Health Sciences of the Norwegian University of Science and Technology. Therapists were recruited prior to randomization such that the allocation was concealed for therapists when they agreed to participate in the study. The therapists were asked to identify consecutive patients without informing the patients of which treatment arm they had been allocated and asked for permission to share contact details with the study team. Patients interested in participating were then contacted by the study team who recruited the patients whilst the allocation was concealed. The CONSORT guidelines do not recommend statistical significance testing at baseline in general, however Bolzern et al (2019) point out that there is greater risk of recruitment bias in cluster randomised trials since the recruitment of individuals occurs post-randomisation. We will therefore estimate the difference in baseline characteristics between the intervention and control groups on an individual level, presenting mean differences with 95 % confidence intervals to identify clinically relevant differences and potential selection bias from the recruitment process.

3.4.2 Blinded statistical analyses

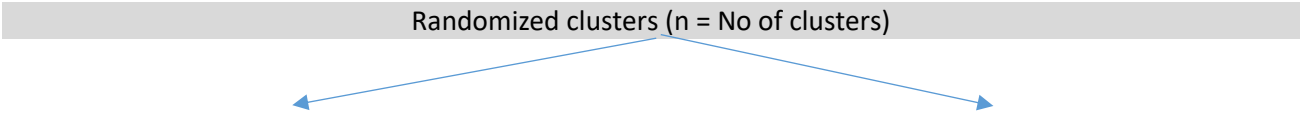
All analyses specified in this document will be performed with the cluster allocation concealed. Two researchers will do the analysis, one blinded who did not participate in the project nor in the data collection. The other one will also be blinded, but has been central in administrating the data collection, and will only work with data where the group allocation is concealed.

4. Presentation of study population

4.1 Screening data, eligibility, recruitment, withdrawal, and follow-up

The screening, eligibility and recruitment processes are described in detail in the protocol. A CONSORT flow diagram will be presented showing how many clinicians were included and how many participants were involved in each step.

As shown in Figure 1, below, the flow chart will present how many patients were invited to the study, how many did not respond, how many declined to participate, how many were excluded and how many were finally included. The flowchart will also present how many patients withdrew from the study and wanted their data deleted, as well as how many responded at each follow-up time-point. Reasons for exclusion, withdrawal, and lost to follow-up will be specified when available.



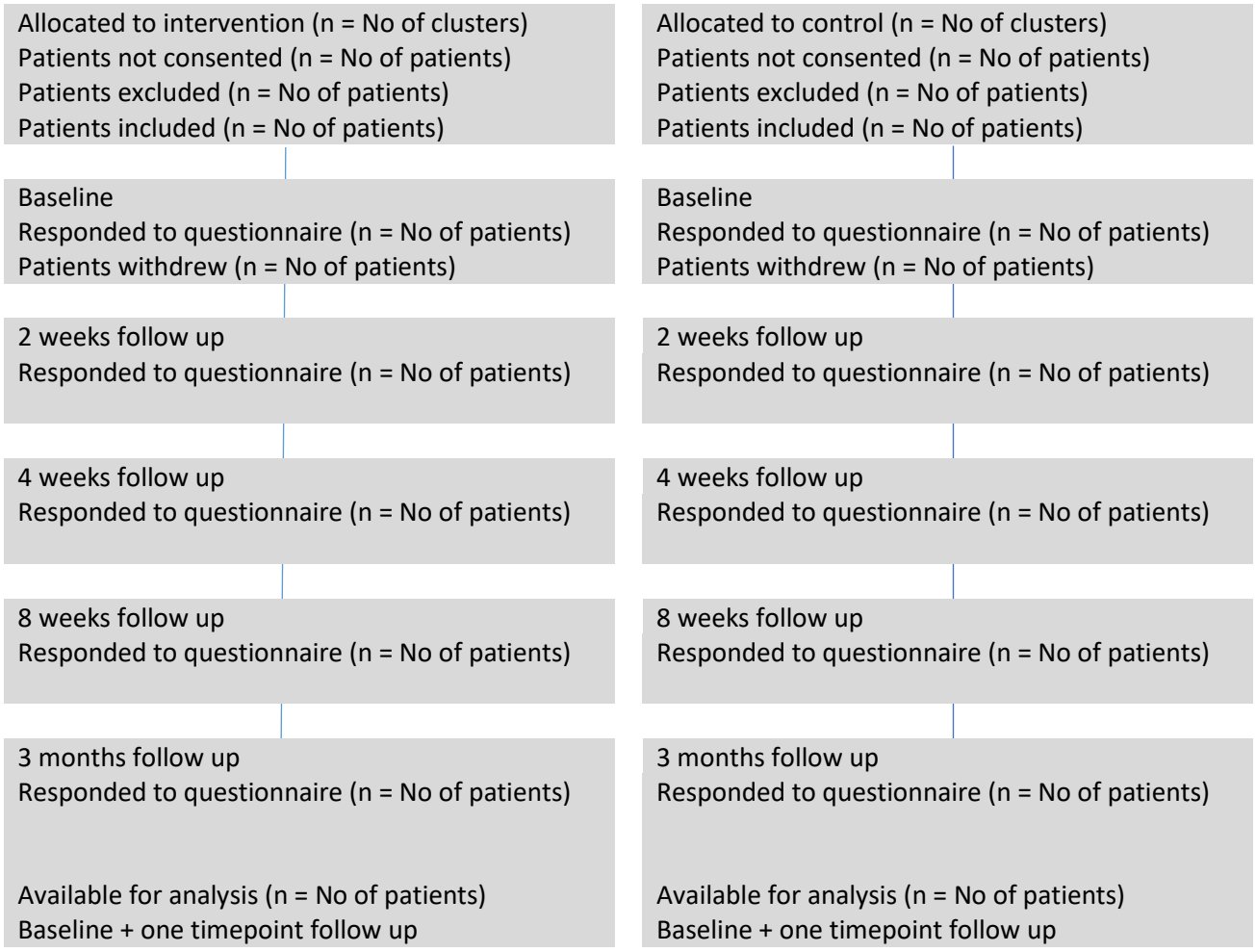


Figure 1: Outline of CONSORT flow diagram.

4.2 Baseline patient characteristics

The following baseline characteristics will be presented for each treatment arm and for the group as a whole.

- Age (years), mean (SD)
- Sex (female), n (%)
- Body mass index (kg/m^2), mean (SD)
- Married or living with partner, n (%)
- Education, n (%): Primary school or less; Highschool; Up to 4 years of higher education; More than 4 years higher education
- Employment status, n (%): Employed; On sick leave; Retired, disability pension, or on work allowance; Student; Other
- Current smoker, n (%)
- Comorbidity, n (%): No comorbidities; 1 comorbidity; 2-3 comorbidities; 4 or more comorbidities
- Pain duration, n (%): Under 1 month; 1- 3 months; 3-6 months; 6-12 months; More than 12 months
- Anxiety for pain in physical activity, mean (SD)
- Diagnosis, n (%): Neck; Shoulder; Back; Hip; Knee; Complex

- The short form Örebro screening questionnaire, mean (SD)
- The MSK Tool, n (%): Low risk; Medium risk; High risk
- Continuous pain, n (%)
- Physical activity- how often do you exercise, n (%): Never; Less than once a week; Once a week; 2-3 times a week; Every day
- Childhood experiences, n (%): Very good; Good; Moderate; Difficult; Very difficult
- Health literacy- difficulty understanding health information, n (%): Never; Rarely; Occasionally; Often; Always

We will also report baseline values for primary and secondary outcomes (see 5.1.1 & 5.2.1). Descriptive statistics will be presented depending on the type and distribution of the variables

5. Analysis

5.1 Analysis of primary outcome

5.1.1 Definition of primary outcome measures

This study has two defined primary outcome measures:

(1) The patient's global perceived effect (GPE) at 3 months after start of treatment measured on a 7-point Likert scale. The GPE scale will be dichotomized as "improved" (score 1-2) or "unchanged/worse" (score 3-7).

(2) The other primary outcome is the proportion with a clinically important improvement at 3 months in function measured by the Patient Specific Function Scale (PSFS; 0-10). An important improvement will be defined as 30% increase on PSFS. Percent changes in PSFS scores will be calculated by taking the absolute change in score divided by the potential change for each individual. For example, a participant who improves from 3 to 5 during the 3-month period will have a 28.6 % improvement (2 of 7 possible points for improvement where 10 is maximum) and this participant would be classified as having not achieved a clinical important improvement.

Both of these outcomes will also be assessed at 2, 4 and 8 weeks, and these will be included in the analysis models for the primary outcomes and presented as secondary analyses, as described below.

5.1.2 Analysis of primary outcomes

The effect of CDSS at 3 months will be estimated for both primary outcomes using three-level mixed logistic regression models. Each primary outcome will be assessed in a separate model that includes the repeated measures of the outcome at 2 weeks, 4 weeks, 8 weeks and 3 months follow-up as the dependent variable, and such that they are clustered by follow-up timepoint (level 1), participants as level 2 and therapists as level 3. Treatment allocation, time point and an interaction between treatment allocation and time point will be included as independent variables. We will adjust for the stratification variable (type of physiotherapist) and possible prognostic variables (age, sex, education and pain-duration). In addition, when analysing PSFS, we will adjust for the baseline value.

The treatment effect will be estimated for each time point from the mixed logistic regression models and presented as an OR with 95 % confidence intervals.

The primary outcome will also be analysed in the per protocol subgroup (§3.3.2) using the same strategy as described above.

5.1.3 Sensitivity analyses

For both primary outcome measures, we will perform a sensitivity analysis including only clusters which managed to recruit at least 10 participants.

We will also do sensitivity analysis for the second primary outcome PSFS, using minimum 3 points improvement from baseline as definition of clinically important improvement at 3 months.

5.1.4 Subgroup analyses and treatment effect heterogeneity

We will do exploratory subgroup analysis to assess for treatment effect heterogeneity for each of the subgroup categorisations defined in §3.3.3. For these analyses, separate mixed logistic regression models will be extended to include each subgroup and an interaction term between treatment allocation and the subgroup. The estimated treatment effect will be estimated and reported for each subgroup. The interaction term will be used for assessing the presence of treatment effect heterogeneity where the comparison is between two subgroups. For the assessment of treatment effect heterogeneity based on the five phenotype groups, overall treatment effect heterogeneity will be assessed based on a likelihood ratio test comparing the models with and without the interaction term. Comparison of treatment effect between phenotype pairs will be considered if the likelihood ratio test indicates an overall treatment effect heterogeneity.

These analyses are considered to be exploratory as the trial was not powered for subgroup analyses. The results from all subgroup analyses will be presented in either the manuscript or supplementary files.

5.1.5 Missing data

The analysis of both primary outcomes will use mixed logistic regression models, incorporating all available data from each participant with at least one follow-up measurement. This method should provide unbiased estimates of the effect of CDSS system under the assumption that the missing data is missing at random (MAR). We have not planned to use any other strategies handling missing data, such as multiple imputations, as this would also depend on the MAR assumption.

Number of participants with missing data for each outcome at each timepoint will be presented as well as a presentation of baseline characteristics of the full sample alongside those included in the primary analysis.

5.2 Analysis of secondary outcomes

5.2.1 Overview of secondary outcomes

Table 1: Overview of secondary outcomes and their planned analyses

Secondary outcomes	Time(s) recorded	Planned analysis	Brief description of recorded data
Pain intensity measured by the Numeric Rating Scale	Baseline, 2 weeks, 8-weeks & 3 months	Linear mixed model	Scored from 0-10, where 0 is no pain
Patient specific functional scale (PSFS)	Baseline, 2 weeks, 4 weeks, 8 weeks & 3 months	Linear mixed model	Scored from 0-10, where 10 is best function
Workability	Baseline, 2 weeks, 4 weeks, 8 weeks & 3 months	Linear mixed model	Scored from 0-10, where 10 is best workability
Pain self-efficacy (PSEQ)	Baseline, 2 weeks, 8 weeks & 3 months	Linear mixed model	Scored from 0-12, where 12 is best
Emotional distress measured by Hopkins Symptom Checklist (HSCL-10)	Baseline & 3 months	Linear regression	Scored from 1.0-4.0, where 1.0 is best
General musculoskeletal health (MSK-HQ)	Baseline & 3 months	Linear regression	Scored from 0-56, where 56 is best health
15D, health-related quality of life. Sleep	Baseline & 3 months	Logistic regression	5 answering options dichotomized into: “I’m able to sleep normally” and “I have slight problems with sleeping” versus “I have moderate problems with sleeping”, “I have great problems with sleeping” and “I suffer severe sleeplessness”.
15D, health-related quality of life. Vitality	Baseline & 3 months	Logistic regression	5 answering options dichotomized into: “I feel healthy and energetic” and “I feel slightly weary, tired or feeble” versus “I feel moderately weary, tired or feeble”, “I feel very weary, tired or feeble, almost exhausted” and “I feel extremely weary, tired or feeble, totally exhausted”
Number of pain sites (pain drawing)	Baseline & 3 months	Linear regression (Poisson to be considered)	From 0-10 pain sites
Use of pain medication last week	Baseline & 3 months	Logistic regression	Yes/No
Number of treatments at 3 months (count)	3 months	Count	Number of physiotherapy treatment sessions up to maximum 3 months
Health related quality of life (EQ-5D)	Baseline & 3 months	Linear regression	Five (5) items each with Likert scale 1 – 5. The answers on the EQ-5D will be transformed into an index value for health status using the UK value set. The index score ranges between − 0.285 (worst imaginable health state) and 1 (perfect health) (Devlin et al., 2018).

5.2.2 Continuous secondary outcomes measured at multiple follow-up timepoints

Continuous secondary outcomes assessed at multiple follow-up timepoints will be assessed using linear mixed models. As for the primary binary outcomes, the repeated measures of each outcome at baseline, 2 weeks, 4 weeks, 8 weeks and 3 months follow-up as the dependent variable, with follow-up timepoint as level 1, participants as level 2 and therapists as level 3. Timepoint and an interaction between treatment allocation and time points will be included as independent variables. Treatment allocation will not be included as a separate main independent variable as recommended by Twisk et al (2018). We will adjust for the stratification variable (type of physiotherapist) and possible prognostic variables (age, sex, education and pain-duration).

The presence of treatment effect heterogeneity in subgroup analysis will be considered for pain intensity (measured on numeric rating scale). As for the two primary outcomes, treatment heterogeneity will be estimated by including an interaction term between treatment allocation and the subgroups defined in §3.3.3, here using separate linear mixed models for each subgrouping.

5.2.3 Continuous/ binary outcomes measured at baseline and one follow-up timepoints

Continuous and binary outcomes measured at baseline and once during the follow-up period will be analysed using linear or logistic regression, respectively. For each outcome, the follow-up measurement will be included as the dependent variable, with treatment allocation as the primary independent variable of interest. These analyses will also be adjusted for the baseline value of the outcome variable, as well as the stratification variable (type of physiotherapist) and possible prognostic variables (age, sex, education and pain-duration). These analyses will include only participants who have completed both the baseline and 3-month follow-up questionnaires. We have not planned any imputation of missing data for these secondary outcomes.

5.2.4 Number of treatments at 3 months

We intend to analyse the number of treatments at 3 months as a continuous variable using a linear regression model equivalent to those described §5.3.4 (without adjustment for a baseline value, as this is not relevant here). The distribution of both the number of treatments and the residuals will be revised to consider if a Poisson regression may be more appropriate given this variable is a count variable. These investigations and a decision about this final strategy will be performed prior to unblinding of the dataset.

5.3 Statistical software

Stata/MP v17.0 will be used to all statistical analyses (College Station, Texas, USA).

6. References

Bolzern, J.E., Mitchell, A. & Torgerson, D.J. Baseline testing in cluster randomised controlled trials: should this be done?. BMC Med Res Methodol 19, 106 (2019). https://doi.org/10.1186/s12874-019-0750-8

Campbell M K, Piaggio G, Elbourne D R, Altman D G. Consort 2010 statement: extension to cluster randomised trials. *BMJ* 2012; 345 :e5661 doi:10.1136/bmj.e5661

Devlin NJ, Shah KK, Feng Y, Mulhern B, van Hout B. Valuing health-related quality of life: an EQ-5D-5L value set for England. *Health Econ.* 2018;27(1):7–22.

Evensen, K.A.I., Robinson, H.S., Meisingset, I. et al. Characteristics, course and outcome of patients receiving physiotherapy in primary health care in Norway: design of a longitudinal observational project. *BMC Health Serv Res*. 2018; 18, 936. https://doi.org/10.1186/s12913-018-3729-y

Gamble C, Krishan A, Stocken D, et al. Guidelines for the Content of Statistical Analysis Plans in Clinical Trials. *JAMA*. 2017;318(23):2337–2343. doi:10.1001/jama.2017.18556

Meisingset I, Vasseljen O, Vøllestad NK, Robinson HS, Woodhouse A, Engebretsen KB, Glette M, Øverås CK, Nordstoga AL, Evensen KAI, Skarpsno ES. Novel approach towards musculoskeletal phenotypes. Eur J Pain. 2020 May;24(5):921-932. doi: 10.1002/ejp.1541. Epub 2020 Feb 27. PMID: 32040225.

Twisk J, Bosman L, Hoekstra T, Rijnhart J, Welten M, Heymans M. Different ways to estimate treatment effects in randomised controlled trials. *Contemp Clin Trials Commun.* 2018 Mar 28;10:80-85. doi: 10.1016/j.conctc.2018.03.008

Wasserstein RL & Lazar NA. The ASA Statement on p-Values: Context, Process, and Purpose. *The American Statistician*. 2016; 70:2, 129-133, DOI: 10.1080/00031305.2016.1154108

Wasserstein RL, Schirm AL & Lazar NA. Moving to a World Beyond "p < 0.05". *The American Statistician*. 2019; 73:sup1, 1-19, DOI: 10.1080/00031305.2019.1583913