STATISTICAL ANALYSIS PLAN (SAP)

An individualised treatment versus a minimal home-based exercise program in women with late-term shoulder impairments after primary breast cancer surgery:

A Statistical Analysis Plan for a Randomised Trial

Authors/Collaborators:

Kim Michéle Feder[1,2,3,4,5], Robin Christensen[4,6], Sabrina Mai Nielsen[4,6], Hans Bjarke Rahr[2,7], Marianne Djernes Lautrup[7,8], Heidi Klakk Egebæk[3,9], and Kim Gordon Ingwersen[1,2,4,5]

Affiliations:

[1] Department of Physiotherapy, Vejle Hospital, University Hospital of Southern Denmark, Beriderbakken 4, DK-7100 Vejle, Denmark.

[2] Department of Regional Health Research, University of Southern Denmark, Campusvej 55, DK-5230 Odense, Denmark.

[3] Research Unit for Applied Health Science, University College South (UC SYD), Lembckesvej 3-7, DK-6100 Haderslev, Denmark.

[4] Section for Biostatistics and Evidence-Based Research, the Parker Institute, Bispebjerg and Frederiksberg Hospital, Nordre Fasanvej 57, DK-2000 Frederiksberg, Denmark.

[5] OPEN – Open Patient data Explorative Network, Odense University Hospital, University of Southern Denmark. J. B. Winsløws Vej 9a, DK-5000 Odense C, Denmark.

[6] Research Unit of Rheumatology, Department of Clinical Research, University of Southern Denmark, Odense University Hospital, Campusvej 55, DK-5230 Odense, Denmark.

[7] Department of Surgery, Vejle Hospital, University Hospital of Southern Denmark, Beriderbakken 4, DK-7100 Vejle, Denmark.

[8] Department of Plastic and Breast Surgery, Aarhus University Hospital, Palle Juul-Jensens Boulevard 35, DK-8200 Aarhus N, Denmark.

[9] Research Unit of Exercise Epidemiology, Institut for Idræt, University of Southern Denmark, Campusvej 55, DK-5230 Odense, Denmark.

Principal Investigator:

Kim Michéle Feder (KMF), Department of Physiotherapy, Vejle Hospital, Vejle - DK and Department of Regional Health Research, University of Southern Denmark, Odense - DK.

Senior Biostatistician and Analysis Support:

Robin Christensen (RC), Professor of Biostatistics and Clinical Epidemiology: Section for Biostatistics and Evidence-Based Research, the Parker Institute, Bispebjerg and Frederiksberg Hospital, Copenhagen, Denmark; and Research Unit of Rheumatology, Department of Clinical Research, University of Southern Denmark, Odense University Hospital, Odense Denmark.

Statistical Analyst:

Sabrina Mai Nielsen (SMN), Section for Biostatistics and Evidence-Based Research, The Parker Institute, Bispebjerg and Frederiksberg Hospital, Denmark & Research Unit of Rheumatology, Department of Clinical Research, University of Southern Denmark, Odense University Hospital, Odense, Denmark.

Kim Michéle Feder, Department of Physiotherapy, Department of Physiotherapy, Vejle Hospital, Vejle - DK and Department of Regional Health Research, University of Southern Denmark, Odense - DK.

Database Manager:

Allan Lind-Thomsen (ALT), Open Patient data Explorative Network (OPEN), Odense University Hospital, Region of Southern Denmark.

Trial Registration Number:

ClinicalTrials.gov (NCT05277909).

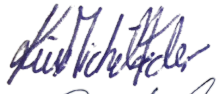
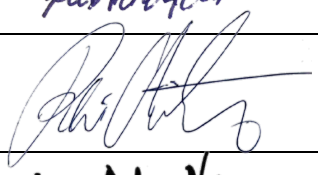
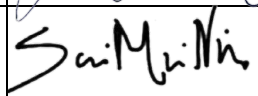
SAP version:

Version 1.0 on August 19, 2022 / October 31, 2022

SAP REVISION HISTORY

Protocol version	SAP version	Section changed in SAP	Description of and reason for change in SAP	Date changed in SAP
3.0	*1.0*	*Sample size and power considerations*	*Lack of recruited patients*	*31.10.2022*

SIGNATURES

Date	Name	Roles and responsibility	Signature
October 31, 2022	PhD-student Kim Michéle Feder	**SAP first author** (Team leader)	
October 31, 2022	Professor Robin Christensen	**Senior Biostatistician** (Risk Manager)	
October 31, 2022	PhD and Postdoc Sabrina Mai Nielsen	**Biostatistician** (Statistical Analyst)	
October 31, 2022	Professor Hans Rahr	**Chief Investigator** (Sponsor)	
October 31, 2022	PhD and Postdoc Kim Gordon Ingwersen	**SAP last author** (Project leader)	

List of abbreviations:

BCS: Breast-Conserving Surgery;

SLND: Sentinel Lymph Node Dissection;

ALND: Axillary Lymph Node Dissection;

QoL: Quality of Life;

Quick DASH: Disabilities of the Arm, Shoulder and Hand;

SPADI: Shoulder Pain and Disability Index;

GPE: Global Perceived Effect;

NRS: Numeric Pain Intensity Rating Scale;

PHQ-9: Patient Health Questionnaire;

GAD-7: General Anxiety Disorder;

OPEN: Open Patient Data Explorative Network;

REDCap: Research Electronic Data Capture;

ITT: Intention-To-Treat;

SD: Standard Deviation;

SE: Standard Error;

CONSORT: CONsolidating Standards Of Reporting Trials;

EQUATOR: Enhancing the QUAlity and Transparency Of health Research network.

SUMMARY

Objectives: The primary efficacy objective is to compare the effect of the individualised treatment plan, relative to a minimal physiotherapeutic rehabilitation program, on changes in Shoulder Pain and Disability Index (SPADI) from baseline to 12 weeks after initiating the treatment, in women with late-term shoulder impairments after primary breast cancer surgery.

Methods: The study was designed as a randomised trial with a two-group comparison testing superiority hypothesis related to the primary and key secondary objectives. In total, 130 participants would randomised (allocation 1:1) to either an expert assessment of shoulder impairments followed by an individualised treatment plan or to follow a minimal physiotherapeutic rehabilitation program delivered in a pamphlet. The primary analysis will be based on the Intention to Treat (ITT) population; the ITT principle asserts the effect of a treatment policy (that is, the planned treatment regimen), rather than the actual treatment given (i.e., it is independent of treatment adherence). The analyses of the key secondary outcomes will be performed in a prioritized order. The analyses of the confirmatory secondary outcomes will be performed in sequence until one of the analyses fails to show the statistically significant difference, or until all analyses have been completed at a two-sided statistical significance level of 0.05. The continuous outcomes will be analysed based on Repeated-Measures Mixed Effects Models (time: 0, 4, 8, and 12 weeks from baseline).

Registration: ClinicalTrials.gov, identifier: NCT05277909. Registered on 11 March 2022.

INTRODUCTION

Breast cancer is the most common cancer in women worldwide. Standard surgical treatment in Denmark is breast-conserving surgery (BCS) or mastectomy in combination with sentinel lymph node dissection (SLND) or axillary lymph node dissection (ALND). In part due to early diagnosis and optimised treatment methods, 5-year survival has improved to currently 87%. Despite fewer mastectomies and more BCS, less invasive surgical procedures of the axilla (e.g., fewer ALND vs. SLND), and more refined radiotherapy procedures, late-term upper limb impairments still remain common after primary breast cancer surgery. The most frequent are lymphoedema, sensory disturbances, pain and impaired shoulder function, with up to 70% of patients reporting at least one of these symptoms three years after surgery. These impairments lead to difficulties in activities of daily living, increased risk of depression and anxiety and decreased quality of life (QoL).

Previous international research has primarily focused on prevention and treatment of lymphoedema, and less on other upper limb impairments. Preoperative and early postoperative physiotherapeutic interventions are known to be effective in reducing shoulder pain and improving shoulder function after breast cancer treatment, but there is a lack of international knowledge on the effectiveness of these interventions on the late-term sequelae.

The aim of this trial is to evaluate whether an expert assessment of shoulder impairments, followed by an individualised treatment plan, is superior to a minimal physiotherapeutic rehabilitation program in reducing shoulder symptoms, among women with late-term shoulder impairments after primary breast cancer.

Rational for this study

Currently, no standardised evaluation or treatment of women's late-term shoulder impairments are offered, and it is therefore up to the individual women to seek care, resulting in large variations in rehabilitation. Since half of all cases of breast cancer is diagnosed in women aged 62 years or younger (1), i.e., potentially physically active and in the workforce, an improvement in prevention and management of shoulder impairment after breast surgery may substantially benefit both the patients and society. This randomised trial will potentially have immediate impact on clinical practice as well as on long-term outcomes and quality of life after breast cancer surgery.

Aims and Hypotheses

The primary aim of this trial is to compare the effectiveness of an individualised treatment, relative to a minimal home-based program, on changes in Shoulder Pain and Disability Index (SPADI) from baseline to week 12 after initiating the treatment, in women with late-term shoulder impairments after breast cancer surgery.

The null hypothesis is that there is no difference in SPADI between the two groups. The alternative hypothesis is that women randomised to the individualised treatment (*Intervention group)* will improve more in shoulder function and reduce shoulder pain 12 weeks after initiating the treatment than those randomised to the minimal home-based program (*Control comparator group)*.

Objectives

Primary efficacy objective:

To compare the effect of the individualised treatment plan, relative to a minimal physiotherapeutic rehabilitation program, on changes in Shoulder Pain and Disability Index (SPADI) from baseline to 12 weeks after initiating the treatment, in women with late-term shoulder impairments after primary breast cancer surgery.

Key secondary objectives:

To compare the effect of the individualised treatment plan, relative to a minimal physiotherapeutic rehabilitation program, after 12 weeks on changes in the following key secondary outcome measures:

- SPADI pain,
- SPADI function,
- SPADI clinical response,
- Impression of the treatment's success,
- Active Range of Motion (A-ROM),
- Passive Range of Motion (P-ROM),
- Number of treatments received due to shoulder symptoms,
- Maximum shoulder pain intensity,

- Shoulder pain during general activities,
- Shoulder pain at rest,
- Shoulder pain during sleep, and
- Shoulder pain during flexion/rotation/abduction.

Exploratory objectives covered will be published in secondary papers covering the following:
To compare the effect of the individualised treatment plan, relative to a minimal physiotherapeutic rehabilitation program, after 12 weeks on changes in:

- Pain medication consumption,
- Depression score (PHQ-9), and
- Anxiety score (GAD-7).

STUDY METHODS

Trial design

This trial is designed as a stratified (by type of surgery and radiotherapy), block randomised (1:1 allocation), controlled, parallel group and assessor-blinded, single-centre superiority trial conducted in Denmark. Women are recruited from an earlier letter inquiry, based on a register extract, including an invitation to participate in a cross-sectional survey of late-term effects. Women reported shoulder impairments invited to participate in this trial and randomised to an expert assessment of shoulder impairments with an individualised treatment plan or to a 12 week home-based program, which is a minimal intervention within the physiotherapeutic rehabilitation and is delivered in a pamphlet. The two interventions are more described specifically in the trial protocol. The primary endpoint will be assessed at the 12 weeks follow-up, while many of the outcome measures are also collected at 4 and 8 weeks follow-up.

The SAP is reported in accordance with the "*Guidelines for the Content of Statistical Analysis Plans in Clinical Trials*" (2). The trial was registered prior to First Patient First Visit at ClinicalTrials.gov (NCT05277909) and follow the principles of the Declaration of Helsinki (3). The recruitment period started in April 2022 (i.e. First Patient First Visit), and expected to be completed in August 2022 (i.e. Last Patient First Visit).

Randomisation, allocation concealment, and blinding

Eligible participants were randomly assigned in permuted blocks of 2 to 6, with a 1:1 allocation, based on a computer-generated randomisation list generated by an independent data manager implemented into the REDCap randomisation system (4), to either *Intervention group* or *Control group*. Participants were stratified in 5 groups according to type of surgery and +/- radiotherapy treatment:

1. Breast Conserving Surgery (BCS) and Sentinel Lymph Node Dissection (SLND) + radiotherapy;
2. BCS and SLND – radiotherapy;
3. BCS and Axillary Lymph Node Dissection (ALND) + radiotherapy;
4. Mastectomy and SLND – radiotherapy; and
5. Mastectomy and ALND + radiotherapy)

Immediately after obtaining the written informed consent, the baseline measures and outcomes will be collected. The outcome assessors who performed the clinical baseline assessment will also perform the 12-week follow-up, blinded towards treatment allocation. Participants were informed about their group allocation directly after randomisation.

To ensure concealment, the primary investigator, assessors and administrators of the randomisation were blinded to the block sizes, as the randomisation code was stored in REDCap, with no access for the project group.

Sample size and power considerations

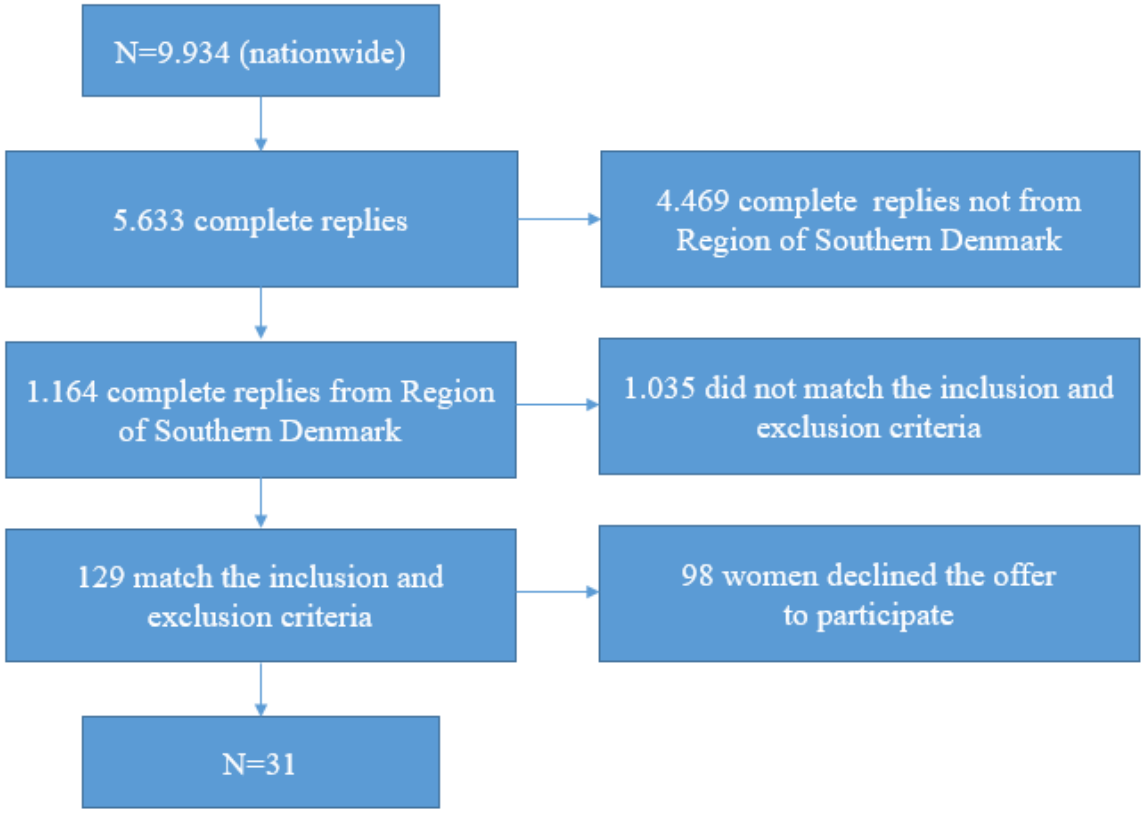
The sample size and power estimation were based on the expected between-group difference in the change in SPADI score from baseline to the 12 weeks follow-up. In order to achieve an *a priori* statistical power of at least 85%, with a two-sided significance level α=0.05, with an anticipated standard deviation (SD)=15.41 SPADI units (5), the estimated total sample size was n=130 (~ 65 participants in each group), to be able to detect a minimal clinically relevant difference defined as 8 points on the SPADI-score (6,7).

Correction:

In the study protocol we estimated an eligible population of approximately n=384 women willing to participate (https://doi.org/10.1186/s13063-022-06659-1), based on the cross sectional survey sent to

approximately 4.300 women in Region of Southern Denmark, corresponding to an expected eligible population rate of 8%.

In fact, we sent out the questionnaire to 9.934 women nationwide, and obtained a participation rate of 56.7% in the cross sectional survey. For the Region of Southern Denmark we received 1.164 complete replies. Of these only 129 women matched the inclusion and exclusion criteria, and furthermore, 76% of these eligible women declined the offer to participate, resulting in n=31, corresponding to a much lower than expected, eligible population rate of 2.66%.



Statistical interim analysis and early stopping rules

No statistical interim analyses were planned or done. Patient recruitment stops when a total number of 130 patients have been included and finished their intervention or when the deadline August 31, 2022 is reached.

Timing of final analysis

The final analysis for the between group comparison (*Intervention group vs. Control comparator group)* for the primary endpoint (baseline, 4-, 8- and 12-weeks follow-up) is planned to be performed after the last randomised patient has completed the 12 weeks follow-up. The publication of the trial will be prepared when the data have been received and cleaned (anticipated by April 2023).

Timing of outcome assessments

The overview of the trial time-points of each outcome assessment is presented in the **Table 1** in the study protocol at the TRIALS journal (*https://doi.org/10.1186/s13063-022-06659-1*): *"Effectiveness of an expert assessment and individualised treatment compared with a minimal home-based exercise program in women with late-term shoulder impairments after primary breast cancer: Study protocol for a randomised controlled trial"*. The intervention start date (*Intervention group* or *Control comparator group*) for each patient is used to calculate the 4-, 8-, and 12-weeks follow-up time points.

STATISTICAL PRICINPLES

Confidence intervals and p-values

All 95% confidence intervals (95% CIs), and *P-values* will be two sided. Adjustments for multiplicity will not be applied, but rather the analyses of the key secondary outcomes will be performed and interpreted in a prioritized order. The analyses of the confirmatory secondary outcomes will be performed in sequence until one of the analyses fails to show the statistically significant difference, or until all analyses have been completed at a statistical significance level of 0.05. The key secondary statistical tests will be reported with *P* values for hypothesis tests and claims of statistical significance.

Analysis populations

The primary analyses will be based on the Intention to Treat (ITT) population, i.e., based on the Full Analysis Set (having the outcome of interest measured at baseline). The ITT principle asserts the effect of a treatment policy (that is, the planned treatment regimen), rather than the actual

treatment given (i.e., it is independent of treatment adherence). Accordingly, participants allocated to a treatment group will be followed up, assessed and analysed as members of that group, irrespective of their adherence to the planned course of treatment (i.e., independent of withdrawals and cross-over phenomena) (8).

TRIAL POPULATION

Screening data

The total number of patients, screened for eligibility, will be collected, and presented in a CONSORT Flow Diagram (**Figure 1**) for documenting the flow of participants through the trial. The number of ineligible patients randomised by mistake, if any, will be reported including reason for ineligibility.

Eligibility criteria

Women who have participated in a parallel performed nationwide cross-sectional survey study and reported shoulder pain or impairment as their primary discomfort after BCS or mastectomy combined with either SLND or ALND due to breast cancer, and fulfilling the following inclusion and exclusion criteria are considered eligible for the randomised trial:

Inclusion criteria

1. Breast cancer patients who underwent unilateral BCS or mastectomy on the left or right side, including SLND or +/- ALND within the last 3-7 years (2015-2019)
2. Currently living in the Region of Southern Denmark or Central Denmark Region with a radius of 75 km from Vejle Hospital
3. Between 18 and 71 age at the time of surgery for primary breast cancer
4. Indicate pain in chest and/or shoulder area (shoulder impairments) as the biggest problem/late-term effect in everyday life
5. Indicate impaired shoulder function due to pain or due to tightness/tension
6. Indicate shoulder pain at rest, during general activities, during sleep or during flexion, rotation or abduction of the shoulder
7. A score ≥15 on the Disabilities of the Arm, Shoulder and Hand (Quick DASH) (9,10)
8. Agree to participate in this trial and sign written informed consent.

Exclusion criteria

1. No previous breast cancer (before 2014)
2. Cancer relapse after the date of index surgery, cancer spread outside of thorax and axilla, tumour fixed to chest wall
3. Primary- or secondary breast reconstruction performed at any time
4. Severe lymphedema (an average score ≥ 70% in the first 7 questionnaires on the LYMPH-ICF-DK (11,12)
5. Bilateral breast cancer surgery
6. Previous surgery in the affected shoulder (prior to inclusion)
7. Previous shoulder or upper limb fractures (left/right)
8. Currently receiving chemo, immuno- or radiotherapy
9. Co-morbidity expected to influence shoulder function (e.g. rheumatoid arthritis, previous stroke, multiple sclerosis)
10. Other reasons for exclusion (e.g. pregnancy, not legally competent, unable to comprehend the information or unable to consent)

Recruitment

The CONSORT Flow Diagram will comprise number of patients screened, excluded (with reasons), randomised, stratified, receiving their allocated treatment, withdrawals (with reasons), and lost to follow-up (with reasons) and included in ITT analysis. The anticipated CONSORT Flow Diagram is depicted in **Figure 1**.

Withdrawal/follow-up

The reason for each crossover or withdrawal will be registered. Participants performing a crossover will remain in the study and be analysed following the intention-to-treat principle (i.e., remaining in their original group). The level of consent withdrawal will be classified as:

1. consent to continue follow-up and data collection;
2. consent to continue data collection only; or
3. complete withdrawal with no further follow-up and data collection (13).

In the CONSORT Flow Diagram numbers and reasons for withdrawal and/or loss to follow-up will be presented given at the 4-, 8- and 12-weeks (primary end point) outcome assessment; separately for the two intervention groups.

Baseline characteristics

The following data will be used to describe patients by randomisation group at baseline: mean age, height, weight, body mass index (BMI), smoking, alcohol consumption, highest education level, employment, index shoulder, dominant side affected, duration of shoulder symptoms, pain medication consumption due to shoulder related pain, depression, anxiety and outcome measures will be collected at baseline. The baseline characteristics are illustrated in **Table 1**.

Categorical variables will be calculated and presented in numbers and percentages. While Means and Standard Deviations (SDs) will be calculated and presented for continuous variables if data follows a normal distribution; in case, continuous variables does not appear to follow normal distribution, median and interquartile range will be calculated. No tests of statistical significance will be conducted for baseline characteristics.

ANALYSIS

Primary outcome

Shoulder Pain and Disability Index (SPADI) [Time Frame: 0, 4, 8 and 12 weeks]

The primary outcome measure will be the between-group difference in change in SPADI-score ($\Delta SPADI_{12}$ = $SPADI12_{\text{-weeks follow-up}}$ - $SPADI_{baseline}$) from baseline to 12 weeks after initiating the treatment. SPADI is considered a valid, reliable, and responsive patient-reported questionnaire assessing shoulder pain and function, where each domain score is equally weighted and added to a total percentage score that ranges from 0 (best) to 100 (worst) (9,14) (**Figure 2**).

Key secondary outcomes

- *SPADI pain [Time Frame: 0, 4, 8 and 12 weeks]*
  A key secondary outcome will be the between-group difference in change in a separate SPADI-pain-subscale ($\Delta SPADIpain_{12}$ = $SPADIpain_{\text{12-weeks follow-up}}$ - $SPADIpain_{baseline}$) from baseline to 12 weeks after initiating the treatment. SPADI is considered a valid and reliable

patient-reported questionnaire consisting of 5-item pain subscale are scored on a numeric rating scale that ranges from 0 (no pain) to 10 (worst pain) (9,14).

- *SPADI function [Time Frame: 0, 4, 8 and 12 weeks]*
  A key secondary outcome will be the between-group difference in change in a separate SPADI-function-subscale ($\Delta SPADIfuncion_{12} = SPADIfuncion_{12\text{-weeks follow-up}} - SPADIfunction_{baseline}$) from baseline to 12 weeks after initiating the treatment. SPADI is a valid and reliable patient-reported questionnaire consisting of 6-item pain subscale (exclusion of question three and seven) are scored on a numeric rating scale that ranges from 0 (no difficulty) to 10 (so difficult that required help) (9,14).

- *SPADI clinical response [Time Frame: 12 weeks (follow-up)]*
  Response to treatment will be computed for the SPADI change score for each woman in both treatment groups and presented dichotomised (i.e. responder and non-responder) as number (and percentages) responders. Women will be classified as a responder if the SPADI change score improves by 8 points or more (≥), corresponding to the minimal clinically important difference on SPADI (6,7) from baseline to 12 weeks follow-up.

- *Global perceived effect (GPE) [Time Frame: 4, 8 and 12 weeks]*
  A key secondary outcome will be the between-group difference in GPE score 12 weeks after initiating the treatment.
  The GPE will evaluate the impression of the treatment's success including overall shoulder problems on a 7-point Likert scale ranging from "markedly worse" to "markedly improved" (15,16).

- *Active Range of Motion (A-ROM) [Time Frame: 0 and 12 weeks]*
  A key secondary outcome will be the between-group difference in change in A-ROM degrees ($\Delta A\text{-}ROM_{12} = A\text{-}ROM_{12\text{-weeks follow-up}} - A\text{-}ROM_{baseline}$) from baseline to 12 weeks after initiating the treatment. The smartphone inclinometer (GetMyROM) is considered valid and reliable to assess A-ROM in flexion, internal rotation, external rotation and abduction (17-21).

- *Passive Range of Motion (P-ROM) [Time Frame: 0 and 12 weeks]*
  A key secondary outcome will be the between-group difference in change in P-ROM degrees (ΔP-ROM$_{12}$ = P-ROM$_{12\text{-weeks follow-up}}$ – P-ROM$_{baseline}$) from baseline to 12 weeks after initiating the treatment. The smartphone inclinometer (GetMyROM) is considered valid and reliable to assess A-ROM in flexion, internal rotation, external rotation and abduction (17-21).

- *Number of treatments received due to shoulder symptoms [Time Frame: 12 weeks (follow-up)]*
  A key secondary outcome will be the between-group difference in mean number of treatments received due to shoulder symptoms from baseline to 12 weeks after initiating the treatment.

- *Maximum shoulder pain intensity [Time Frame: 0, 4, 8 and 12 weeks]*
  A key secondary outcome will be the between-group difference in change in maximum pain intensity (ΔMaxPain$_{12}$ = MaxPain$_{12\text{-weeks follow-up}}$ – MaxPain$_{baseline}$) from baseline to 12 weeks after initiating the treatment. The Numeric Rating Scale (NRS) is considered a valid, reliable and responsible single 11-item patient-reported pain scale that ranges from 0 (no pain) to 10 (worst pain) (22).

- *Shoulder pain during general activities [Time Frame: 0, 4, 8 and 12 weeks]*
  A key secondary outcome will be the between-group difference in change in shoulder pain during general activities (ΔGeneral activity pain$_{12}$ = General activity pain$_{12\text{-weeks follow-up}}$ – General activity pain$_{baseline}$) from baseline to 12 weeks after initiating the treatment. The NRS is considered a valid, reliable and responsible single 11-item patient-reported pain scale that ranges from 0 (no pain) to 10 (worst pain) (22).

- *Shoulder pain at rest [Time Frame: 0, 4, 8 and 12 weeks]*
  A key secondary outcome will be the between-group difference in change in shoulder pain at rest (ΔPain at rest$_{12}$ = Pain at rest$_{12\text{-weeks follow-up}}$ – Pain at rest$_{baseline}$) from baseline to 12 weeks after initiating the treatment. The NRS is considered a valid, reliable and responsible single

11-item patient-reported pain scale that ranges from 0 (no pain) to 10 (worst pain) (22).

- *Shoulder pain during sleep [Time Frame: 0, 4, 8 and 12 weeks]*
  A key secondary outcome will be the between-group difference in change in shoulder pain during sleep (ΔPain during sleep$_{12}$ = Pain during sleep$_{12\text{-weeks follow-up}}$ – Pain during sleep$_{baseline}$) from baseline to 12 weeks after initiating the treatment. The NRS is a valid, reliable and responsible single 11-item patient-reported pain scale that ranges from 0 (no pain) to 10 (worst pain) (22).

- *Shoulder pain during flexion/rotation/abduction [Time Frame: 0 and 12 weeks]*
  A key secondary outcome will be the between-group difference in change in shoulder pain during flexion/rotation/abduction (ΔPain during flex rot abd$_{12}$ = Pain during flex rot abd$_{12\text{-weeks follow-up}}$ – Pain during flex rot abd$_{baseline}$) from baseline to 12 weeks after initiating the treatment. The NRS is a valid, reliable and responsible single 11-item patient-reported pain scale that ranges from 0 (no pain) to 10 (worst pain) (22).

The primary and key secondary outcomes will be presented as illustrated in **Table 2.**

Analysis methods

The CONSORT guideline (13) will be followed for all trial reporting aspects, as recommended by the *"Enhancing the QUAlity and Transparency Of health Research"* (EQUATOR) network website (23).

The primary (continuous) outcome will be analysed by using repeated measures mixed effects models, including participants as a random effect, with fixed effect factors for group (2 levels) and week (4 levels for the SPADI questionnaire [weeks 0, 4, 8, and 12]), the stratification factors and the interaction between group and week, adjusted for baseline values. To assess the adequacy of the linear models describing the observed data—and checking assumptions for the systematic and the random parts of the models— the model features will be investigated via the predicted values and the studentized residuals; that is, the residuals have to be normally distributed (around 0) and be independent of the predicted values. Results will be expressed based on least squares mean estimates with standard errors (SE) as well as the differences in the least squares means with 95% CIs to represent precision of the contrast between groups. Further, for the primary

outcome, a 95% CI excluding differences greater than 4 SPADI points between groups will be interpreted as indicating absence of a clinically meaningful difference. Dichotomous outcome variables will be analysed with logistic regression, with identical fixed effect factors and covariates as the mixed linear model described above. Missing data for dichotomous outcomes will be computed based on conservative (non-responder) imputations.

Subgroup analyses (24) will be used to examine whether the observed overall treatment effect varies across participants subgroups, and to whether the effect is modified by the value of a variable assessed at baseline, analysed by the following thresholds: median age, median duration of shoulder symptoms, obesity (BMI ≥30 kg/m^2), dominant side affected (left *vs* right). This statistical approach to evaluate potential effect modifiers, will be a test for statistical interaction whether the treatment effect (net benefit on change in the SPADI score) varies across levels of the effect modifier (25).

Missing data and sensitivity analyses

To handle missing data, repeated-measures linear mixed effects models will be used (26-29). These models are considered valid if data are "*Missing at Random*" (i.e., any systematic difference between the missing values and the observed values can be explained by differences in observed data) (30).

The following four strategies for handling and interpretation of missing data will be applied in the sensitivity analysis:

1. We will attempt to follow up all randomised participants, even if they withdraw from allocated treatment (i.e., contact all individuals unless they explicitly stated that they had withdrawn their consent);

2. Perform a main analysis of all observed data that are valid under a plausible assumption about the missing data (i.e., data as randomised; using linear mixed effects models, assuming that data are *"Missing at Random"*);

3. Perform sensitivity analyses to explore the effect of departures from the assumption made in the main analysis (i.e., non-responder-imputation: using the value at baseline to replace missing data;

these models will be informative about the robustness in case data are "*Missing Not At Random*"); and

4. Account for all randomised participants, at least in the sensitivity analyses (including strategy 2 and 3, plus the corresponding analyses based on the per protocol population) (28).

Thus, the following sensitivity analyses will be performed on various population analyses, including a non-responder imputation and per-protocol, to examine the robustness by revealed similar results in these sensitivity analyses. Robustness is a concept that refers to the sensitivity of the overall conclusions to various limitations of the data, assumptions, and analytic approaches to data analysis. Robustness implies that the treatment effect and primary conclusions of the trial are not substantially affected when analyses are carried out based on alternative assumptions or analytic approaches.

Additional analyse

No additional analyses on the primary and key secondary outcomes are planned from baseline to 12 weeks follow-up.

Harms

Serious adverse events (SAE) will be categorised in accordance with the definitions established by the United States Food and Drug Administration (e.g. hospitalisation or death) (31). All categorical types of SAE´s will be presented for each group and statistical tested by using logistic regression. The SAEs will be presented as illustrated in **Table 3.**

Statistical software

All data analysis will follow the statistical methods described above, and will be performed applying STATA (Statacorp, College Station, Texas, USA) and SAS (SAS Institute Inc., Cary, North Carolina, USA) software.

REFERENCES

1. American Cancer Society. Breast Cancer Facts & Figures 2019-2020 [Internet]. 2020 [cited 2022 Aug 08, pp. 4]. Available from: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/breast-cancer-facts-and-figures/breast-cancer-facts-and-figures-2019-2020.pdf
2. Gamble C, Krishan A, Stocken D, et al. *Guidelines for the Content of Statistical Analysis Plans in Clinical Trials*. Jama 2017;**318**(23):2337-43.
3. World Medical Association (WMA). WMA Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects [Internet]. 1964 [cited 2018 Aug 14]. Available from: https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects
4. Harris P.A., Taylor R., Thielke R. et al. *Research electronic data capture (REDCap) -- a metadata-driven methodology and workflow process for providing translational research informatics support.* J Biomed Inform. 2009;**42**(2):377–81.
5. Walters S.J. et al. *Sample size estimation for randomised controlled trials with repeated assessment of patient-reported outcomes: what correlation between baseline and follow-up outcomes should we assume?* Trials. 2019;**20**(1):566.
6. Roy J.S., MacDermid J.C., and Woodhouse L.J. *Measuring Shoulder Function: A Systematic Review of Four Questionnaires.* Arthritis Rheum. 2009:**61**(5):623-632.
7. Breckenridge J.D., and McAuley J.H. *Shoulder Pain and Disability Indes (SPADI).* J Physiother. 2011;**57**(3):197.
8. Detry MA, Lewis RJ. *The intention-to-treat principle: how to assess the true effect of choosing a medical treatment*. Jama 2014;**312**(1):85-6.
9. Angst F., Schwyzer H.K., Aeschlimann A., Simmen B.R., and Goldhahn J. *Measures of Adult Shoulder Function. Disabilities of the Arm, Shoulder, and Hand Questionnaire (DASH) and Its Short Version (QuickDASH), Shoulder Pain and Disability Index (SPADI), American Shoulder and Elbow Surgeons (ASES) Society Standardized Shoulder Assessment Form, Constant (Murley) Score (CS), Simple Shoulder Test (SST), Oxford Shoulder Score (OSS), Shoulder Disability Questionnaire (SDQ), and Western Ontario Shoulder Instability Index (WOSI).* Arthritis Care & Research. 2011;**63**(11):174–188.

10. Schønnemann J.O. and Eggers J. *Validation of the Danish version of the Quick-Disabilities of Arm, Shoulder and Hand Questionnaire.* Dan Med J. 2016;**63**(12):A5306.
11. Grarup K.R., Devoogdt N., and Strand L.V. *The Danish version of Lymphoedema Functioning, Disability and Health Questionnaire (Lymph-ICF) for breast cancer survivors: Translation and cultural adaptation followed by validity and reliability testing.* Physiotherapy Theory and Practice. 2019;**35**(4):327-340.
12. Devoogdt N., Van Kampen M., Geraerts I., Coremans T., and Christiaens M.-R. *Lymphoedema Functioning, Disability and Health Questionnaire (Lymph-ICF): Reliability and Validity*. Phys Ther. 2011;**91**(6):944–957.
13. Moher D., Hopewell S., Schulz K.F., Montori V., Gøtzsche P.C., Devereaux P.J. et al. *CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials*. BMJ. 2010;**23**(3):340:c869.
14. Christiansen D.H., Andersen J.H., and Haahr J.P. *Cross-cultural adaption and measurement properties of the Danish version of the Shoulder Pain and Disability Index.* Clinical Rehabilitation. 2012;**27**(4):355-360.
15. Kamper S.J., Maher C.G., and Mackay G., *Global rating of change scales: a review of strengths and weaknesses and considerations for design.* J Man Manip Ther. 2009;**17**(3):163-70.
16. Ramiro S. et al. *The OMERACT Core Domain Set for Clinical Trials of Shoulder Disorders.* The Journal of Rheumatology. 2019;**46**(8):696-75.
17. Mejia-Hernandez K. et al. *Smartphone applications for the evaluation of pathologic shoulder range of motion and shoulder scores – a comparative study*. JSES Open Access. 2018;**2**(1):109-114.
18. Mitchell K. et al. *Reliability and validity of goniometric iPhone applications for the assessment of active shoulder external rotation*. Physiother Theory Pract. 2014;**30**(7):521-525.
19. Keogh J. W. L. et al. *Reliability and validity of clinically accessible smartphone applications to measure joint range of motion: a sysyematic review*. PLoS One. 2019;**14**(5):1-24.
20. Werner BC et al. *Validation of an innovative method of shoulder range-of-motion measurement using a smartphone clinometer application.* J Shoulder Elbow Surg. 2014;**23**:275-282.

21. Shin S. H. et al. *Within-day reliability of shoulder range of motion measurement with a smartphone. Manual Therapy.* 2012;**17**(4):298-304.
22. Paice J.A., Cohen F.L. *Validity of a verbally administered numeric rating scale to measure cancer pain intensity.* Cancer Nurs. 1997;**20**(2):88-93.
23. Christensen R., Bliddal H. and Henriksen M. *Enhancing the reporting and transparency of rheumatology research: a guide to reporting guidelines.* Arthritis Res Ther. 2013:28;**15**(1):109.
24. Wang R, Lagakos SW, Ware JH, Hunter DJ and Drazen JM. *Statistics in Medicine – Reporting of Subgroup Analyses in Clinical Trials*. N Engl J Med 2007;**357**(21):2189-94.
25. Christensen R, Bours MJL, Nielsen SM. *Effect Modifiers and Statistical Tests for Interaction in Randomized Trials.* J Clin Epidemiol 2021;**134**:174-77.
26. Dmitrienko A. and D´Agostino RB. *Multiplicity Considerations in Clinical Trials*. N Engl J Med 2018;**378**(22):2115-22.
27. Little RJ, D'Agostino R, Cohen ML, et al. *The prevention and treatment of missing data in clinical trials.* N Engl J Med 2012;**367**(14):1355-60.
28. White IR, Horton NJ, Carpenter J, et al. *Strategy for intention to treat analysis in randomized trials with missing outcome data.* Bmj 2011;**342**:d40.
29. Ware JH, Harrington D, Hunter DJ and D'Agostino RB. *Missing Data.* N Engl J Med 2012;**367**(14):1353-1354.
30. Detry MA, Ma Y. *Analyzing repeated measurements using mixed models*. *Jama* 2016;**315**(4):407-8.
31. Administration, U.S.F.D. What is a Serious Adverse Event? [Internet]. 2016 [cited 2019 26 Oct]. Available from: https://www.fda.gov/safety/medwatch/howtoreport/ucm053087.html

ANTICIPATED OUTLINE

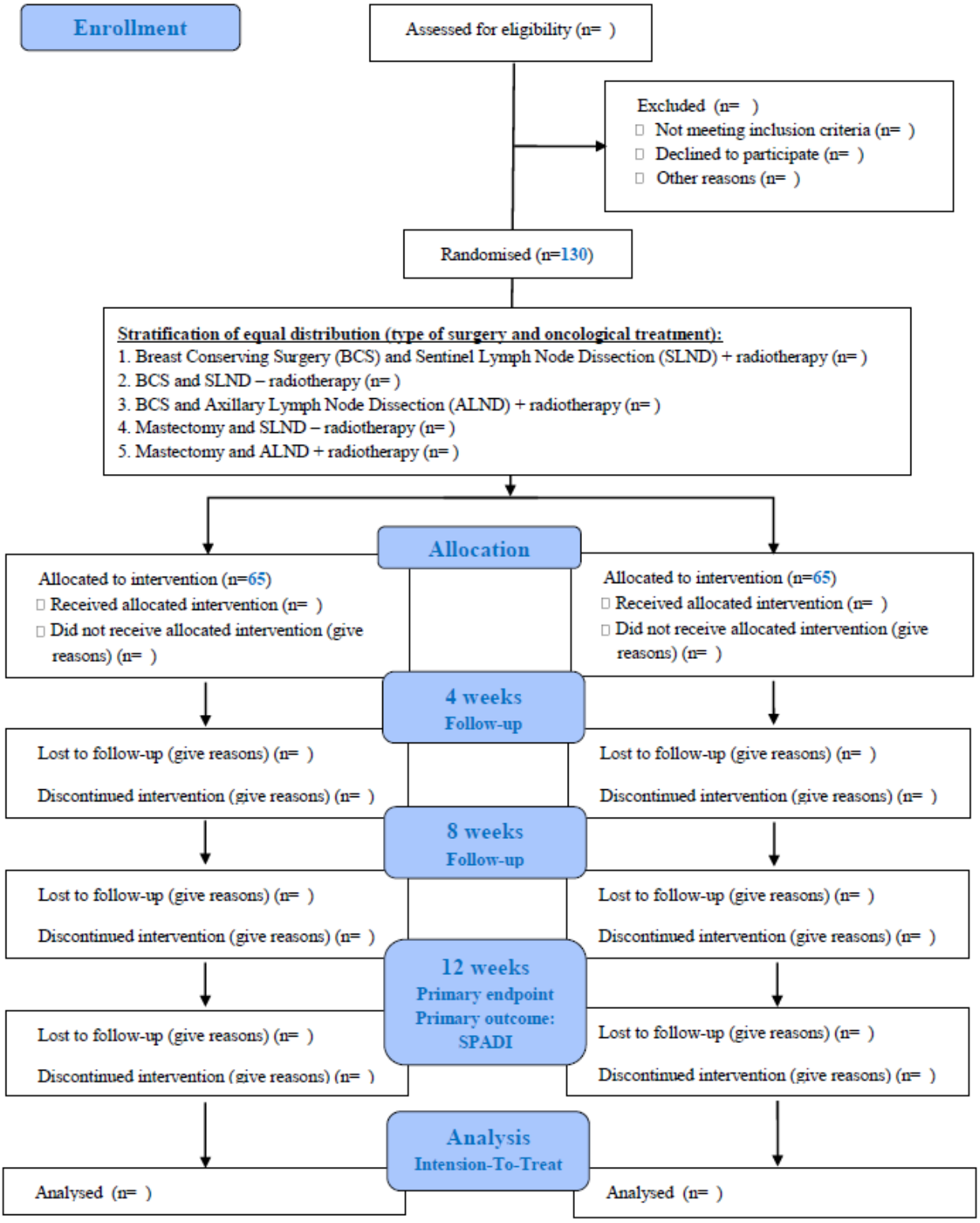


Figure 1: CONSORT Flow Diagram of participants through the trial. *Control comparator group (a minimal home-based program); Intervention group (individualised treatment).*

THE FIGURE BELOW IS DEPICTED WITH ANTICIPATED CHANGES FROM BASELINE TO 12 WEEKS FOLLOW-UP!

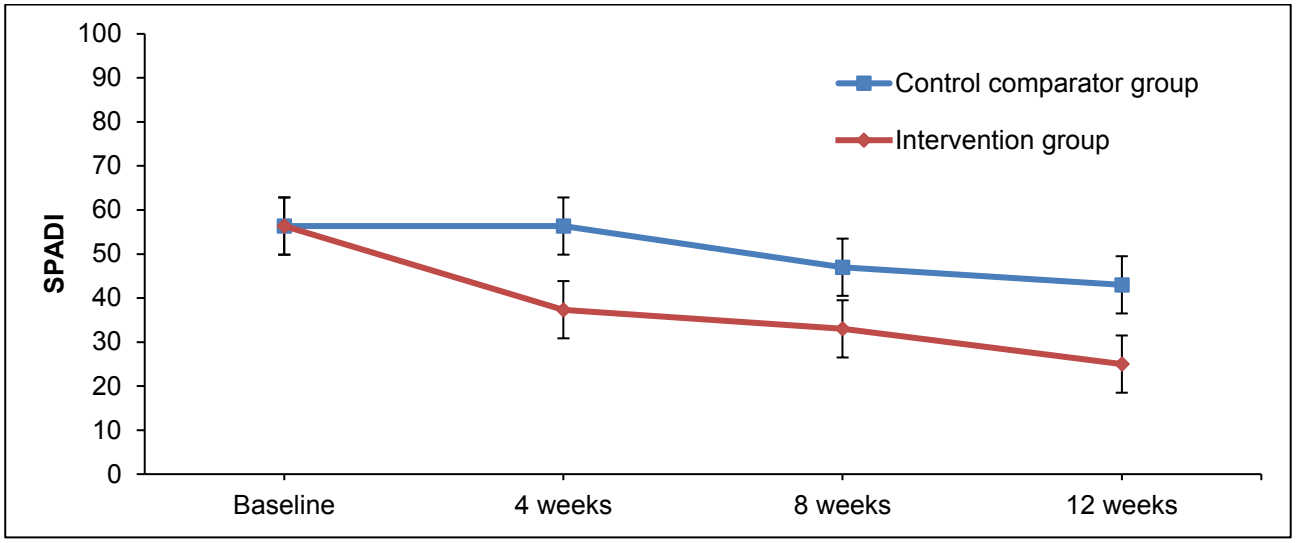


Figure 2: Trajectory in SPADI-score in the *Control comparator group (a minimal home-based program)* and *Intervention group (individualised treatment)* at baseline, 4-, 8-, and 12 (primary end-point) weeks after initiating the treatment. Values are mean (95% CI).

Table 1. Baseline characteristics in the ITT population*

	***Intervention Group* (N =XX)**	***Control Comparator Group* (N =XX)**	**Total Combined (N =XX)**
General characteristics			
Mean age — years			
Height — cm			
Weight — kg			
Body Mass Index — kg/m^2			
Smoking Habits:			
Smoker— no. (%)			
Current smoker— no. (%)			
Not a smoker— no. (%)			
Highest Education level:			
Short — no. (%)			
Medium — no. (%)			
Long — no. (%)			
Employment:			
Employed for wages — no. (%)			
Self-employed — no. (%)			
Sick leave — no. (%)			
Retired — no. (%)			
Other — no. (%)			
Index shoulder:			
Right side — no. (%)			
Left side — no. (%)			
Dominant side affected:			
Right side — no. (%)			
Left side — no. (%)			
Duration of shoulder symptoms — years			
Outcome measures			
SPADI score† — 0 to 100			
Shoulder pain — 0 to 10			
Shoulder function — 0 to 10			
GPE impression of the treatment´s success — 0 to 7			
A-ROM in the affected shoulder — degree			
P-ROM in the affected shoulder — degree			
NRS maximum shoulder pain intensity — 0 to 10			
NRS shoulder pain during general activities — 0 to 10			
NRS shoulder pain at rest — 0 to 10			
NRS shoulder pain during sleep — 0 to 10			
NRS shoulder pain assessment during flexion/rotation/abduction — 0 to 10			

* Plus–minus values are mean ±SD unless otherwise indicated

† The Shoulder Pain and Disability Index (SPADI) ranges from 0 (best) to 100 (worst), with lower scores indicating better disease status.

Table 2: Primary and Key Secondary Outcomes at 12 weeks in the ITT population*

Outcome	***12 weeks after initiating the treatment***				***Between-Group Difference in Mean Improvement***	
	Intervention Group		*Control Comparator Group*		*Difference in LSMeans (95%CI)*	*P-Value*
	Mean	SE	Mean	SE		
Primary endpoint						
Change SPADI score — 0 to 100						
Key secondary outcome measures						
Change SPADI pain (0-10)						
Change SPADI function (0-10)						
Change Clinical response (SPADI change score)						
GPE impression of the treatment success (0-7)						
Change A-ROM in the affected shoulder (degree)						
Change P-ROM in the affected shoulder (degree)						
Number of treatments received due to shoulder symptoms (mean)						
NRS maximum shoulder pain intensity (0-10)						
NRS shoulder pain during general activities (0-10)						
NRS shoulder pain at rest (0-10)						
NRS shoulder pain during sleep (0-10)						
NRS shoulder pain assessment during flexion/ rotation/abduction (0-10)						
Response to Treatment						
SPADI minimal important change criteria† no. (%)						

* All analyses will be based on the Intention-To-Treat (ITT) population: Using Repeated-Measures, Mixed Models (with no simple imputation for missing data); Estimates will be least squares means (LSMeans) and Standard Errors (SE) with the difference between groups reported with 95% Confidence Intervals (CI).

† Patients will be classified as having a clinically relevant change if the SPADI score improves by 8 points or more; missing data will be replaced by single-step non responder imputation.

Table 3: Serious Adverse Events at 12 weeks in the ITT population*

	Intervention Group	*Control Comparator Group*
Serious Adverse Event (SAE) - no (%)		
Musculoskeletal		
Deep infection, no. (%)		
Shoulder dislocation, no. (%)		
Shoulder fracture, no. (%)		
Cardiovascular		
Vascular injury (arterial and venous damage), no. (%)		
Pulmonary embolism, no. (%)		
Deep venous thrombosis, no. (%)		
Acute myocardial infarction, no. (%)		
Nervous system		
Nerve injury, no. (%)		
Deaths, no. (%)		
Discontinuation due to AE(s) — no. (%)		

* This table includes all Serious Adverse Events (SAE´s) that occurred during the 12 weeks study period, but not necessarily a causal relationship with the treatment administered. An adverse event was classified as serious if it was fatal or life-threatening, required or prolonged inpatient hospitalisation, was disabling, resulted in (a congenital anomaly or birth defect), or required medical or surgical intervention to prevent permanent impairment or damage.

Appendix Table 1: Sensitivity Analysis at 12 weeks in the ITT population using a single step non-responder imputation*

Outcome	***12 weeks after initiating the treatment***		***Between-Group Difference in Mean Improvement***	
	Intervention Group	*Control Comparator Group*	*Difference in LSMeans (95%CI)*	*P-Value*
Primary endpoint				
SPADI score — 0 to 100				
Key secondary outcome measures				
SPADI pain (0-10)				
SPADI function (0-10)				
Clinical response (SPADI change score)				
GPE impression of the treatment success (0-7)				
A-ROM in the affected shoulder (degree)				
P-ROM in the affected shoulder (degree)				
Number of treatments received due to shoulder symptoms (mean)				
NRS maximum shoulder pain intensity (0-10)				
NRS shoulder pain during general activities (0-10)				
NRS shoulder pain at rest (0-10)				
NRS shoulder pain during sleep (0-10)				
NRS shoulder pain assessment during flexion/ rotation/abduction (0-10)				

* The primary outcome will be analysed using single step imputation, based on the data set where missing data is handled. All key secondary outcomes are analysed based on the Repeated-Measures, Mixed Models (based on all collected time points): Missing data will be replaced by simple single step imputation, where missing data is replaced by the value at baseline (baseline observation carried forward).

Appendix Table 2: Sensitivity Analysis at 12 weeks in the *Per Protocol* population*

Outcome	***12 weeks after initiating the treatment***		***Between-Group Difference in Mean Improvement***	
	Intervention Group†	*Control Comparator Group‡*	*Difference in LSMeans (95%CI)*	*P-Value*
Primary endpoint				
SPADI score — 0 to 100				
Key secondary outcome measures				
SPADI pain (0-10)				
SPADI function (0-10)				
Clinical response (SPADI change score)				
GPE impression of the treatment success (0-7)				
A-ROM in the affected shoulder (degree)				
P-ROM in the affected shoulder (degree)				
Number of treatments received due to shoulder symptoms (mean)				
NRS maximum shoulder pain intensity (0-10)				
NRS shoulder pain during general activities (0-10)				
NRS shoulder pain at rest (0-10)				
NRS shoulder pain during sleep (0-10)				
NRS shoulder pain assessment during flexion/ rotation/abduction (0-10)				

* All analyses will be based on the *Per Protocol* population: Using Repeated-Measures, Mixed Models (with no imputation for missing data); Estimates are least squares means (LSMeans) and Standard Errors (SE) with difference between groups reported with 95% Confidence Intervals (CI).

† The Per-Protocol Population included N=?? in the Intervention Group.

‡ The Per-Protocol Population included N=?? in the Control Comparator Group.

Appendix Table 3: Sensitivity Analysis at 12 weeks in the *As-Treated* population*

Outcome	*Intervention Group*†	*Control Comparator Group*‡	*Intervention Group and no expert assessments*§	*Control Comparator Group and expert assessments*¶	*P-Value*
Primary endpoint					
SPADI score — 0 to 100					
Key secondary outcome measures					
SPADI pain (0-10)					
SPADI function (0-10)					
Clinical response (SPADI change score)					
GPE impression of the treatment success (0-7)					
A-ROM in the affected shoulder (degree)					
P-ROM in the affected shoulder (degree)					
Number of treatments received due to shoulder symptoms (mean)					
NRS maximum shoulder pain intensity (0-10)					
NRS shoulder pain during general activities (0-10)					
NRS shoulder pain at rest (0-10)					
NRS shoulder pain during sleep (0-10)					
NRS shoulder pain assessment during flexion/ rotation/abduction (0-10)					

* All analyses will be based on the *As-Treated* population: Using Repeated-Measures, Mixed Models; Estimates are least squares means and Standard Errors (SE). An As-Treated analysis will be performed based on their patients' adherence to the randomised treatment expecting four groups:

† Patients randomised to the Intervention Group undergoing an expert assessment included N=??.

‡ Patients randomised to the Control Comparator Group without undergoing expert assessments in the follow-up period included N=??.

§ Patients randomised to the Intervention Group but declined an expert assessment post randomisation included N=??.

¶ Patients randomised to the Control Comparator Group undergoing expert assessments during the follow-up period N=??.