Statistical Analysis Plan

TRIAL FULL TITLE	Feasibility of a Positive Affect Skills Intervention for Reducing Racial Disparities in Pain Management
TRIAL REGISTRATION	NCT04869345
SAP VERSION	1.0
SAP VERSION DATE	February 21, 2023
TRIAL STATISTICIAN	Kenneth Tyler Wilcox, PhD, Cornell University
Protocol Version (SAP associated with)	1.0
TRIAL PRINCIPAL INVESTIGATOR	Anthony Ong, PhD, Cornell University Cary Reid, MD PhD, Weill Medical College of Cornell University
SAP AUTHOR(s)	Kenneth Tyler Wilcox, PhD, Cornell University

1 SAP Signatures

I give my approval for the attached SAP entitled Feasibility of a Positive Affect Skills Intervention for Reducing Racial Disparities in Pain Management dated February 21, 2023.

Statistician (Author)
Name: Kenneth Tyler Wilcox, PhD

Signature: 

Date: February 21, 2023

Principal Investigator
Name: Anthony Ong, PhD

Signature:

2/21/23

Date:

2 Table of Contents

1 SAP Signatures 1
2 Table of Contents 2
3 Abbreviations and Definitions 4
4 Introduction 4
4.1 Preface 4
4.2 Scope of the analyses 5
5 Study Objectives and Endpoints 5
5.1 Study Objectives 5
5.1.1 Objective 5
5.1.2 Hypotheses / Research Questions 6
5.2 Endpoints 6
5.2.1 Primary Endpoints 6
5.2.2 Secondary Endpoints 6
6 Study Methods 6
6.1 General Study Design and Plan 6
6.2 Inclusion-Exclusion Criteria and General Study Population 7
6.2.1 Study Population 7
6.2.2 Inclusion Criteria 7
6.2.3 Exclusion Criteria 7
6.3 Randomization and Blinding 7
6.4 Study Assessments 8
Table 1. Schedule of trial events 8
6.4.1 Primary Outcomes 8
6.4.1.1 Recruitment as Measured by Rates of Enrollment 8
6.4.1.2 Retention as Measured by Change in Enrollment 8
6.4.1.3 Helpfulness, Satisfaction, and Impact as Assessed by Self-Report Participant Feedback Survey 9
6.4.1.4 Length of Intervention Time as Measured by Number of Weeks to Complete the Intervention 9
6.4.2 Secondary Outcomes 9
6.4.2.1 Patient-Reported Outcomes Measurement Information System (PROMIS) Pain Measures9
6.4.2.2 Change in Depressive Symptoms as Measured by the Center for Epidemiologic Studies Depression Scale Revised (CESD-R-10) 10
6.4.2.3 Change in Stress Appraisal as Measured by the Perceived Stress Scale 10

6.4.2.4 Change in Positive Affect as Measured by the Modified Differential Emotions Scale (mDES) 10
6.4.2.5 Change in Positive Affect as Measured by the Positive and Negative Affect Scale (PANAS-GEN) 10
6.4.2.6 Daily Survey Assessments 10
7 Sample Size 11
8 General Analysis Considerations 11
8.1 Analysis Populations 11
8.1.1 Modified Intention to Treat Population 11
8.2 Covariates and Subgroups 11
8.3 Missing Data 12
8.4 Multiple Testing 12
9 Summary of Study Data 12
9.1 Subject Disposition 12
Table 2. Study CONSORT flow diagram 14
Baseline 14
Analysis 14
Post-Intervention 14
1-Month Follow-Up 14
Enrollment 14
9.2 Demographic and Baseline Variables 15
9.2.1 Demographic Variables 15
9.2.1.1 Age 15
9.2.1.2 Sex 15
9.2.1.3 Ethnicity 15
9.2.1.4 Race 15
9.2.2 Baseline Variables 15
9.2.2.1 PROMIS Pain Intensity Short Form 3a 15
9.2.2.2 PROMIS Pain Interference Short Form 6b 15
9.2.2.3 PROMIS Physical Functioning Short Form 10a 15
9.2.2.4 PROMIS Fatigue Short Form 6a 15
9.2.2.5 Change in Depressive Symptoms as Measured by the Center for Epidemiologic Studies Depression Scale Revised (CESD-R-10) 15
9.2.2.6 Change in Stress Appraisal as Measured by the Perceived Stress Scale 15
9.2.2.7 Change in Positive Affect as Measured by the Modified Differential Emotions Scale (mDES) 15

9.2.2.8 Change in Positive Affect as Measured by the Positive and Negative Affect Scale (PANAS-GEN) 15
9.2.2.9 Change in Affective Reactivity to Stress as Measured by the Daily Modified Differential Emotions Scale (mDES) 15
9.2.2.10 Change in Affective Reactivity to Stress as Measured by the Daily Inventory of Stressful Events (DISE) 15
9.3 Concurrent Illnesses and Medical Conditions 15
10 Efficacy Analyses 16
10.1 Primary Efficacy Analysis 16
10.2 Secondary Efficacy Analyses 17
10.2.1 Analyses of Secondary Endpoints 17
11 Safety Analyses 17
11.1 Extent of Exposure 17
11.2 Adverse Events 17
11.3 Deaths, Serious Adverse Events, and other Significant Adverse Events 18
11.4 Pregnancies 18
11.5 Prior and Concurrent Medications 18
12 Reporting Conventions 18
13 References 18

3 Abbreviations and Definitions

AE	Adverse Event
CRF	Case Report Form
CTSC	Clinical Translational Science Center
FMS	Fibromyalgia Syndrome
LARKSPUR	Lessons in Affect Regulation to Keep Stress and Pain UndeR control
PA	Positive Affect
SAP	Statistical Analysis Plan
REDCap	Research Electronic Data Capture

4 Introduction

4.1 Preface

Chronic non-cancer pain affects as many as 100 million Americans and is associated with significant functional limitations and physical disability. Standard behavioral therapies typically focus on minimizingnegative thoughts and emotions associated with pain and yield only modest treatment effects. Efforts aretherefore needed to develop more effective

psychological treatments for chronic pain by identifying new targets for intervention. We propose to pilot test a resilience-based intervention that targets positive affect (PA) in fibromyalgia syndrome (FMS) patients, a chronic pain population with known deficits in PA and an inability to regulate PA in the face of pain. Our goal is to conduct a randomized pilot trial of LARKSPUR (Lessons in Affect Regulation to Keep Stress and Pain UndeR control), an online-delivered PA skills intervention. Our central hypothesis is that our LARKSPUR program will (a) show acceptability and feasibility in engaging and retaining FMS patients and (b) demonstrate greater improvements in PA and FMS-related pain and functional impairment.

This proposal for pilot data collection will lay the foundation for a high-quality randomized trial for people with FMS. The proposed work holds promise as an effective, low cost, scalable, and readily implementable novel non-pharmacologic intervention to help people cope with FMS. By demonstrating the feasibility, acceptability, and preliminary efficacy of the LARKSPUR intervention, we hope to improve quality of life, as well as the options available for care for millions of people who suffer from FMS.

4.2 Scope of the analyses

Aim 1. Feasibility and acceptability will be examined by conducting frequency and descriptive statistics (i.e., mean, median, standard deviation, range) for enrollment rates, number of sessions completed, number of weeks required to complete the intervention, and Likert-scale items assessing satisfaction withthe intervention and perceived helpfulness.
Aim 2. Aim 2 will be addressed using a pre-post design. To determine the degree to which the LARKSPUR intervention is likely to improve patient outcomes, a post-versus-pre-difference will be sought to estimate the change in patients' outcomes (e.g., pain self-management). This is a small feasibility trial to explore the acceptability and feasibility of the intervention among racially and ethnically diverse older adults with fibromyalgia. The trial is not powered to estimatestatistically reliable differences in outcomes.

5 Study Objectives and Endpoints

5.1 Study Objectives

Our goal for this study is to evaluate the feasibility, acceptability, and effect size of a previously developed online positive affect (PA) skills intervention—LARKSPUR (Lessons in Affect Regulation to Keep Stress and Pain UndeR control)—in a sample of patients with fibromyalgia syndrome (FMS).

5.1.1 Objective

5.1.1.1 To maximize relevance and acceptability of content and delivery of LARKSPUR intervention among patients with FMS, a chronic pain population with known deficits in PA. This aim will establish the feasibility (recruitment and retention) and acceptability (helpfulness, satisfaction, and impact) of the multicomponent LARKSPUR intervention in patients with FMS.
5.1.1.2 To conduct a randomized pilot trial to estimate the effect size of the LARKSPUR intervention in FMS pain (primary outcome), as well PA, depressive symptoms, physical functioning, and stress appraisals (secondary outcomes) and explore racial/ethnic disparities. We hypothesizethat intervention participants will report more frequent PA, decreased depressive symptoms,enhanced physical functioning, improved stress appraisals,

and reduced FMS pain (intensity and interference) immediately following the intervention (approximately 8 weeks), and at 1-Month Follow-Up.

5.1.2 Hypotheses / Research Questions

We hypothesize that the intervention will lead to more frequent daily PA, which, in addition to directly affecting depressive symptoms, will also have indirect effects through enhanced physicalfunctioning and improved stress appraisals (*resilience mechanisms*). Decreased depressive symptoms, enhanced physical functioning, and improved stress appraisals, in turn, are hypothesized to lead to reduced FMS pain (*resilience outcomes*).

5.2 Endpoints

5.2.1 Primary Endpoints

5.2.1.1 Recruitment as Measured by Rates of Enrollment
5.2.1.2 Retention as Measured by Change in Enrollment
5.2.1.3 Helpfulness, Satisfaction, and Impact as Assessed by Self-Report Participant Feedback Survey
5.2.1.4 Length of Intervention Time as Measured by Number of Weeks to Complete Intervention

5.2.2 Secondary Endpoints

5.2.2.1 Change in FMS Pain as Measured by PROMIS Pain Intensity – Short Form 3a
5.2.2.2 Change in FMS Pain as Measured by PROMIS Pain Interference – Short Form 6b
5.2.2.3 Change in Depressive Symptoms as Measured by the Center for Epidemiologic Studies Depression Scale Revised (CESD-R-10)
5.2.2.4 Change in Positive Affect as Measured by the Modified Differential Emotions Scale (mDES)
5.2.2.5 Change in Positive Affect as Measured by the Positive and Negative Affect Scale (PANAS-GEN)
5.2.2.6 Change in Physical Functioning as Measured by PROMIS Physical Functioning Short Form 10a
5.2.2.7 Change in Physical Functioning as Measured by PROMIS Fatigue – Short Form 6a
5.2.2.8 Change in Stress Appraisal as Measured by the Perceived Stress Scale
5.2.2.9 Change in Affective Reactivity to Stress as Measured by the Daily Modified Differential Emotions Scale (mDES)
5.2.2.10 Change in Affective Reactivity to Stress as Measured by the Daily Inventory of Stressful Events (DISE)

6 Study Methods

6.1 General Study Design and Plan

The trial is a randomized, longitudinal, parallel-group, attention-controlled trial. Treatment allocation is a 1:1 ratio. Participants are randomly assigned to LARKSPUR or an emotion-reporting-only attention control condition stratified by race/ethnicity; non-Hispanic Black, Hispanic, or non-Hispanic other (e.g., White, Asian) after completion of eligibility screening

and baseline assessments (see Table 1 for the schedule of trial events). All participants will know their intervention/attention control status.

The LARKSPUR study protocol states that one of the Primary Outcomes is to establish feasibility (Retention as Measured by Change in Enrollment), so Retention will be tested for superiority.

The LARKSPUR study protocol states that the secondary objectives include hypotheses that "intervention participants will report more frequent PA, decreased depressive symptoms, enhanced physical functioning, improved stress appraisals, and reduced FMS pain (intensity and interference) immediately following the intervention (approximately 8 weeks), and at 1-Month Follow-Up." Therefore, all Secondary Endpoints will be tested for superiority.

Other Primary Outcomes characterize the treatment group only and are not tested statistically.

6.2 Inclusion-Exclusion Criteria and General Study Population

6.2.1 Study Population

Participants 50 years of age or older with a fibromyalgia diagnosis that will be stratified by race/ethnicity; non-Hispanic Black, Hispanic, or non-Hispanic other (e.g., White, Asian).

6.2.2 Inclusion Criteria

6.2.2.1 Access to daily internet
6.2.2.2 Male or female 50 years of age or older
6.2.2.3 Fluent in English and able to read and write in English
6.2.2.4 Physician diagnosis of FMS AND/OR Score ≥13 on the 6-item, self-report fibromyalgia screening tool
6.2.2.5 Report having pain for at least the last three months

6.2.3 Exclusion Criteria

6.2.3.1 Cognitive impairment
6.2.3.2 Current behavioral treatment for pain
6.2.3.3 Enrolled in another pain study

6.3 Randomization and Blinding

Participants are randomized to LARKSPUR or an emotion-reporting-only attention control condition stratified by race; non-Hispanic Black, Hispanic, or non-Hispanic other (e.g., White, Asian).

The random number sequence will be generated and uploaded into REDCap by non-data collecting study staff. Allocation concealment will be utilized to prevent selection bias; group assignment will be given to both the participant and selected study staff only after completion of the baseline assessment.

6.4 Study Assessments

Table 1. Schedule of trial events

Table 1: Type and Timing of Measures	PS	B	R	DS	PI	M1
Screening						
Age ≥ 50, read and understand English, FMS diagnosis, 6-item self-report	●					
FMS screening tool, cognitive functioning, no other behavioral treatment forpain, internet access, race/ethnicity	●					
Demographic and Clinical Characteristics						
Age, race/ethnicity, sex, income, education, employment, marital status		●				
Charlson Comorbidity Index		●				
Pain						
Standardized Pain Measures		●			●	●
Daily report of pain and fatigue				●		
Physical Function						
Body Mass Index (BMI)		●				
PROMIS Physical Function Short Form 10a		●			●	●
PROMIS Fatigue Short Form 6a		●			●	●
Exercise in the last 24 hrs		●		●	●	●
Psychosocial Factors						
Standardized Psychosocial Factor Measures		●			●	●
Daily stressors (DISE)				●		
Daily positive and negative affect				●		
Daily positive events				●		
Feedback and Care Updates						
Feedback Survey					●	
Self-Enrolled Therapy Update						●

PS = pre-study; B = Baseline; R = randomization; DS = daily survey; PI = Post-Intervention; M1 = 1-Month Follow-Up

6.4.1 Primary Outcomes

6.4.1.1 Recruitment as Measured by Rates of Enrollment

Enrollment is assessed based on the percentage of participants recruited into the study out of all total eligible participants.

% Enrollment = (number of recruited participants meeting eligibility requirements and enrolled into the study) / (number of recruited participants) * 100.

6.4.1.2 Retention as Measured by Change in Enrollment

Retention is assessed based on the percentage of participants enrolled at Baseline that completed the Post-Intervention assessment.

% Retention Post-Intervention = (number of participants enrolled at Baseline that completed the Post-Intervention assessment) / (number of participants enrolled at Baseline) * 100.

% Retention 1-Month Follow-Up = (number of participants enrolled at Baseline that completed the Post-Intervention assessment and 1-Month Follow-Up assessment) / (number of participants enrolled at Baseline that completed the Post-Intervention assessment) * 100.

6.4.1.3 Helpfulness, Satisfaction, and Impact as Assessed by Self-Report Participant Feedback Survey

6.4.1.3.1 Rating of Recommending Intervention to Friend

The item is measured on an 11-point Likert scale for which 0=Definitely Not to 10=Definitely Yes. Assessed at Post-Intervention.

6.4.1.3.2 Rating of Recommending Intervention to Those with Chronic Pain

The item is measured on an 11-point Likert scale for which 0=Definitely Not to 10=Definitely Yes. Assessed at Post-Intervention.

6.4.1.4 Length of Intervention Time as Measured by Number of Weeks to Complete the Intervention

Length of Intervention Time is assessed based on the number of weeks from the time participants first accessed the online intervention platform to the time participants completed all unique weekly home practice exercises at least once. Assessed at Post-Intervention.

6.4.2 Secondary Outcomes

6.4.2.1 Patient-Reported Outcomes Measurement Information System (PROMIS) Pain Measures

Raw scores are converted to T-scores are derived using item response theory scoring by HealthMeasures <https://healthmeasures.net>. Assessed at Baseline, Post-Intervention, and 1-Month Follow-Up.

6.4.2.1.1 PROMIS Pain Intensity Short Form 3a

This 3-item instrument assesses how much a person hurts [5]. The first two items assess pain intensity utilizing a 7-day recall period (items include the phrase "the past 7 days") while the last item asks patients to rate their pain intensity "right now." Respondents report their pain on a 5-point scale: 1=Had no pain, 2=Mild, 3=Moderate, 4=Severe, and 5=Very severe. PROMIS Pain Intensity T-scores range from 36.3 (had no pain) to 81.8 (very severe pain).

6.4.2.1.2 PROMIS Pain Interference Short Form 6b

This 6-item instrument measures the self-reported impact of pain on a person's life and extent to which pain may interfere with engagement with social, cognitive, emotional, physical, and recreational activities over a 7-day recall period [6]. Respondents report levels of pain interference on a 5-point scale: 1=Not at all, 2=A little bit, 3=Somewhat, 4=Quite a bit, and 5=Very much. PROMIS Pain Interference T-scores range from 41.0 (no pain interference) to 78.3 (highest pain interference).

6.4.2.1.3 PROMIS Physical Functioning Short Form 10a

This 10-item instrument measures a patient's abilities and limitations with respect to everyday physical activities like climbing stairs, carrying groceries, and being able to sit on/get up from the toilet [7]. Respondents report limitations on a 5-point scale: 5=Not at all, ..., 3=Somewhat, ..., 1=Cannot do, and abilities to perform activities on a 5-point scale: 5=Without any difficulty, ..., 3=With some difficulty, ..., 1=Unable to do. PROMIS Physical Functioning T-scores range from 13.5 (limited physical function) to 61.9 (high physical function).

6.4.2.1.4 PROMIS Fatigue Short Form 6a

This 6-item instrument assesses a patient's level of fatigue over a 7-day recall period [8].

Respondents are asked to report fatigue on a scale on a five-point scale: 1=Not at all, 2=A little bit, 3=Somewhat, 4=Quite a bit, 5=Very much. PROMIS Fatigue is scored on the T-score metric (higher scores indicate greater pain intensity).

6.4.2.2 Change in Depressive Symptoms as Measured by the Center for Epidemiologic Studies Depression Scale Revised (CESD-R-10)

The 10-item version of the Center for Epidemiological Studies Depression Scale (CES-D) asks for participants to rate how often over the past week they experienced symptoms associated with depression on a 4-point scale [9]: 0=Rarely or None of the time, 1=Some or Little of the time, 2=Moderately or Much of the time, 3=Most or Almost all the time. CESD-R-10 mean scores range from 0 (rare depressive symptoms) to 4 (depressive symptoms most or almost all the time). Assessed at Baseline, Post-Intervention, and 1-Month Follow-Up.

6.4.2.3 Change in Stress Appraisal as Measured by the Perceived Stress Scale

The 10-item version of the Perceived Stress Scale (PSS) assesses the perception of stress over the previous month [10]. Respondents report how often they have experienced perceived stress on a five-point scale: 0=Never, 1=Almost never, 2=Sometimes, 3=Fairly often, 4=Very often. PSS mean scores range from 0 (experienced no or minimal stress) to 4 (experienced very frequent stress). Assessed at Baseline, Post-Intervention, and 1-Month Follow-Up.

6.4.2.4 Change in Positive Affect as Measured by the Modified Differential Emotions Scale (mDES)

The modified Differential Emotions Scale (mDES) is a 20-item instrument that measures the extent to which a patient has experienced positive and negative emotions over a chosen time frame; in the version used in this study, we ask for emotions over the past 7 days [11]. Respondents are asked to report the greatest amount of a given emotion on a five-point scale: 0=Not at all, 1=A little bit, 2=Moderately, 3=Quite a bit, 4=Extremely. mDES mean scores range from 0 (not at all experienced positive emotion) to 4 (extremely positive emotion). Assessed at Baseline, Post-Intervention, and 1-Month Follow-Up.

6.4.2.5 Change in Positive Affect as Measured by the Positive and Negative Affect Scale (PANAS-GEN)

The 20-item self-report Positive and Negative Affect Scale (PANAS-GEN) asks participants to describe to what extent they feel different feelings and emotions on average [12]. Respondents answer on a five-point scale: 0=Very slightly or not at all, 1=A little, 2=Moderately, 3=Quite a bit, 4=Extremely. PANAS-GEN mean scores range from 0 (very slightly or not at all positive affect) to 4 (extremely positive affect). Assessed at Baseline, Post-Intervention, and 1-Month Follow-Up.

6.4.2.6 Daily Survey Assessments

Assessed over 7-day periods at Baseline, Post-Intervention, and 1-Month Follow-Up.

6.4.2.6.1 Change in Affective Reactivity to Stress as Measured by the Daily Modified Differential Emotions Scale (mDES)

A 20-item instrument that measures the extent to which a patient has experienced positive and negative emotions over a chosen time frame; in the version used in this survey, we ask for emotions over the past 24 hours [11]. Respondents are asked to report the greatest amount of a given emotion on a five-point scale: 0=Not at all, 1=A little bit, 2=Moderately, 3=Quite a bit, 4=Extremely. mDES mean scores range from 0 (not at all experienced positive emotion) to 4 (extremely positive emotion).

6.4.2.6.2 Change in Affective Reactivity to Stress as Measured by the Daily Inventory of Stressful Events (DISE)

A 7-item self-report instrument in which participants report whether stressful events have occurred within the past 24 hours, indicating 1=yes or 0=no accordingly [13].

7 Sample Size

Given that the primary purpose of this study is to assess the feasibility of methods and procedures to be used in a larger, fully powered trial of LARKSPUR, sample size was chosen based on feasibility indicators (recruitment, retention, adherence) rather than on formal power calculations, as appropriate for the pilot nature of the study [see 1]. For quantitative pilot studies, sample sizes of 30 per treatment arm have been proposed to establish feasibility [2]. To allow for attrition between end-of-trial and follow-up, we aimed to recruit 45 per group (intervention and control). Findings from subgroup analyses should therefore be interpreted with caution and call for further confirmatory randomized trials [3].

8 General Analysis Considerations

8.1 Analysis Populations

8.1.1 Modified Intention to Treat Population

All subjects who received any study drug and who participated in at least one post-baseline assessment (i.e., all randomized participants with data following the treatment intervention window).

Participants may be removed from the Modified Intention to Treat Population for analysis if they have (a) no observed data on Post-Intervention assessments, (b) withdrew consent during the study, or (c) were unable to access the online intervention and daily emotion reporting platform.

8.2 Covariates and Subgroups

Baseline covariates for each secondary outcome are expected to correlate with their matching secondary outcome scores at Post-Intervention and 1-Month Follow-Up assessments. Therefore, analyses of each secondary outcome will include participant Baseline scores for adjustment of superiority analyses of treatment differences.

Race/ethnicity was used to stratify treatment allocation and will be adjusted for in primary analyses if sample sizes within each race/ethnicity—treatment subgroup is sufficient (i.e., greater than 5). Otherwise, race/ethnicity will be omitted from primary analyses.

Given the pilot nature of the study, sample size was chosen based on feasibility indicators (recruitment, retention, adherence) rather than on formal power calculations [see 1]. Accordingly, interaction analyses comparing treatment effects by race/ethnicity, age and gender are exploratory.

8.3 Missing Data

All missing data due to early discontinuation of participants from the study or otherwise will be handled using mixed-effects modeling (i.e., multilevel modeling) with participant random intercepts. Multilevel models yield unbiased and consistent estimates of model parameters of interest in the presence of missing outcomes under either Missing Completely at Random or Missing at Random missing data mechanisms. Sensitivity to this assumption will be assessed by comparing analytic results using the modified intention to treat sample to complete-case analyses. Complete-case analyses yield unbiased and consistent estimates under a Missing Completely at Random missing data mechanism, although they are less efficient due to the removal of participant data.

Study personnel will make concerted efforts to follow-up with participants that prematurely discontinue participation in the study to understand potential reasons for their discontinuation where possible.

Participant retention and attrition will be summarized at each primary assessment point in the study (Baseline, Post-Intervention, and 1-Month Follow-Up).

8.4 Multiple Testing

Because this is a pilot study for the purpose of establishing feasibility, control of the overall study-wide significance level is of lesser concern. However, it is important to reflect the statistical uncertainty in analytic results. Accordingly, analytic results will present confidence intervals alongside treatment effect estimates and p-values where appropriate. Kenward-Roger adjustments to the corresponding standard error estimates, degrees of freedom, confidence intervals, and p-values will be used for the multilevel model analyses of secondary outcomes to control the significance level [4].

9 Summary of Study Data

Feasibility and acceptability will be examined by conducting frequency and descriptive statistics (i.e., non-missing sample size, mean, median, standard deviation, maximum, minimum) for enrollment rates, number of sessions completed, number of weeks required to complete the intervention, and Likert-scale items assessing satisfaction with the intervention and perceived helpfulness. Generally, all data will be summarized by treatment and visit number within subject. All summary tables will be structured with a column for each treatment in the order (Control, Experimental) and will be annotated with the total population size relevant to that table/treatment, including any missing observations. Note that the sample size of non-missing values for univariate summary statistics may be larger than the sample size of non-missing values in a complete-case analysis, as different participants may have missing values in different variables, such as baseline covariates and the endpoints.

9.1 Subject Disposition

Enrollment at Baseline and Retention at Post-Intervention and 1-month Follow-Up will be established as described in Section 6.4.

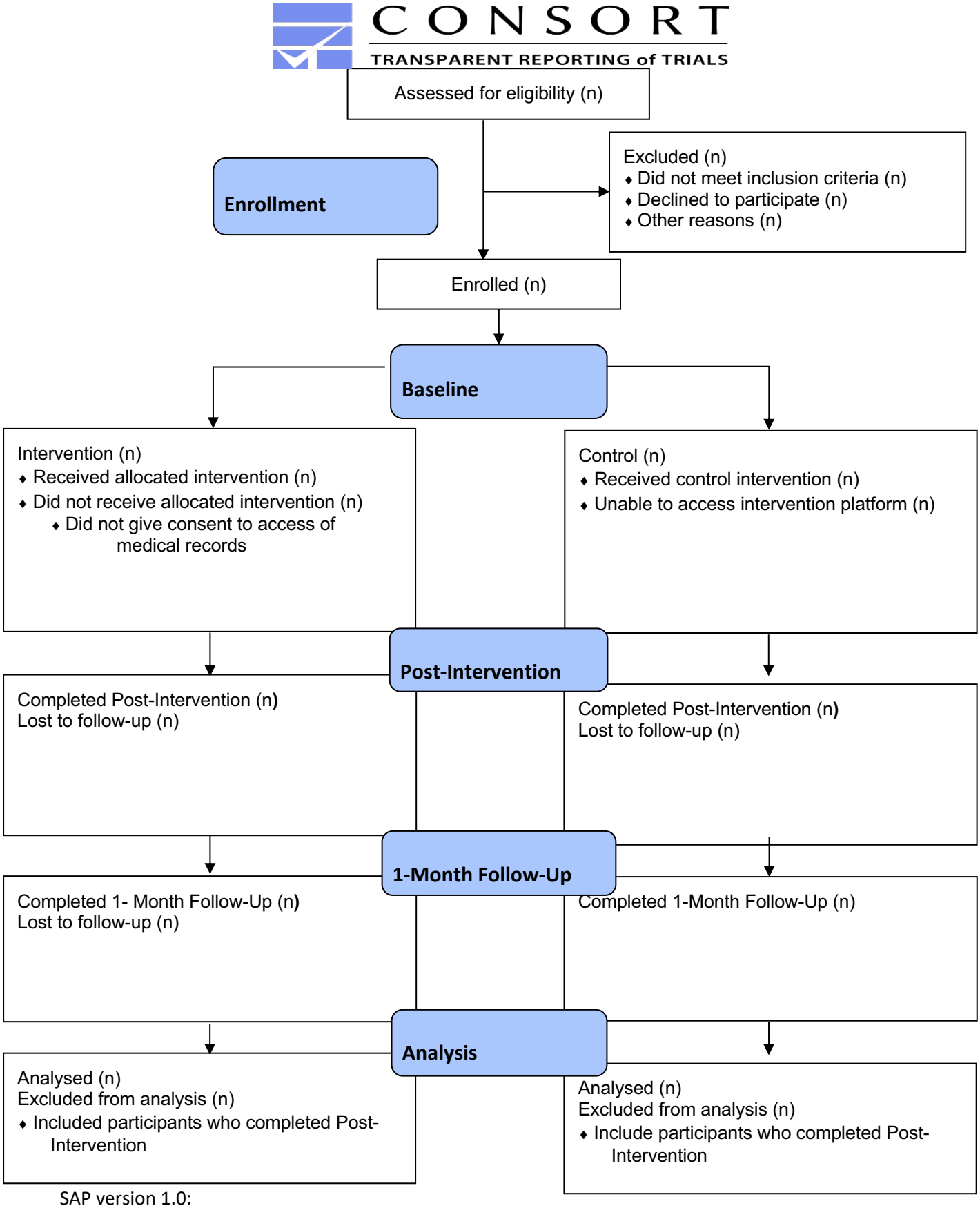
Reasons for participant exclusion during enrollment will include failure to meet inclusion criteria (see Section 6.2), unwillingness to provide informed consent, etc.

Reasons for participant attrition may include withdrawal of consent, disrupted technological access, non-responsiveness during the intervention period on the intervention platform, or failure to complete Post-Intervention or 1-Month Follow-Up assessments.

Participant disposition summary statistics will be produced in accordance with Section 9.

A skeleton CONSORT flow diagram is provided below.

Table 2. Study CONSORT flow diagram



9.2 Demographic and Baseline Variables

9.2.1 Demographic Variables

9.2.1.1 Age

50-59 years, 60-69 years, 70-79 years, ≥ 80 years)

9.2.1.2 Sex

Male, Female

9.2.1.3 Ethnicity

NIH/OMB Categories: Hispanic or Latino, Not Hispanic or Latino, Unknown or Not Reported

9.2.1.4 Race

NIH/OMB Categories: American Indian or Alaska Native, Asian, Native Hawaiian or Other Pacific Islander, Black or African American, White, More than one race, Unknown or Not Reported

9.2.2 Baseline Variables

9.2.2.1 PROMIS Pain Intensity Short Form 3a

9.2.2.2 PROMIS Pain Interference Short Form 6b

9.2.2.3 PROMIS Physical Functioning Short Form 10a

9.2.2.4 PROMIS Fatigue Short Form 6a

9.2.2.5 Change in Depressive Symptoms as Measured by the Center for Epidemiologic Studies Depression Scale Revised (CESD-R-10)

9.2.2.6 Change in Stress Appraisal as Measured by the Perceived Stress Scale

9.2.2.7 Change in Positive Affect as Measured by the Modified Differential Emotions Scale (mDES)

9.2.2.8 Change in Positive Affect as Measured by the Positive and Negative Affect Scale (PANAS-GEN)

9.2.2.9 Change in Affective Reactivity to Stress as Measured by the Daily Modified Differential Emotions Scale (mDES)

9.2.2.10 Change in Affective Reactivity to Stress as Measured by the Daily Inventory of Stressful Events (DISE)

The summary statistics will be produced in accordance with Section 9.

9.3 Concurrent Illnesses and Medical Conditions

Chronic comorbid medical conditions will be assessed using an interviewer-administered version of the Charlson Comorbidity Index [14].

10 Efficacy Analyses

All efficacy variables will be summarized by treatment group in accordance with Section 9.

Except for the two Daily Survey Measures [Change in Affective Reactivity to Stress as Measured by the Daily Modified Differential Emotions Scale (mDES) and Change in Affective Reactivity to Stress as Measured by the Daily Inventory of Stressful Events (DISE)], all Secondary Outcomes will be analyzed using mixed-effects analysis of variance with treatment group by time (Post-Intervention, 1-Month Follow-Up) adjusting for the mean-centered Baseline scores on the respective secondary outcome. Models will include fixed effects for treatment (LARKSPUR vs. control), time (Post-Intervention, 1-month Follow-Up), and the interaction between treatment and time. Random intercepts for participants will be included to account for within-participant correlation over time. The modified Intention-to-Treat sample described in Section 8.1 will be used. Treatment groups will be compared at Post-Intervention, at 1-Month Follow-Up, and with respect to their change from Post-Intervention to 1-Month Follow-Up using 2-sided tests at the 5% significance level under the null hypothesis of no difference in means/change. All assumptions for the mixed-effects models will be assessed by inspection of plots of within-participant and between-person residuals. Standard errors, degrees of freedom, 95% confidence intervals, and p-values will be adjusted using the Kenward-Roger method to control the significance level [4]. Restricted maximum likelihood (REML) estimation will be used. All analyses are exploratory with the aim of hypothesis generation as described in the Study Protocol given the pilot nature of this study.

10.1 Primary Efficacy Analysis

The Recruitment Primary Outcome will be calculated as described in Section 6.4.

The Retention Primary Outcome treatment group percentages will be compared using Boschloo's Test at the 2-sided 5% significance level under the null hypothesis that retention rates are equal between treatment groups [15]. The maximum likelihood estimate of the odds ratio of treatment retention to control retention will be computed with a 2-sided 95% confidence interval and p-value.

The Rating of Recommending Intervention to Friend and Rating of Recommending Intervention to Those with Chronic Pain items will be summarized using an ordinal regression (proportional odds) model without predictors (i.e., the ordinal properties of the Likert scales of these two items will be respected by estimating the average rating adjusted for the cumulative category response probability distribution) [16]. The modified Intention-to-Treat sample described in Section 8.1 will be used.

The Length of Intervention Time will be calculated as described in Section 6.4.

10.2 Secondary Efficacy Analyses

Exploratory interaction analyses along demographic (ethnicity, race) and clinical covariates will assess treatment effect heterogeneity for the primary outcome analyses in Section 10.1 by summarizing/testing Primary Outcomes by subgroup.

10.2.1 Analyses of Secondary Endpoints

Secondary Outcomes will be analyzed as described in Section 10 for the eight survey assessment measures detailed in Section 6.4.2.

The Change in Affective Reactivity to Stress as Measured by the Daily Modified Differential Emotions Scale (mDES) measure and the Change in Affective Reactivity to Stress as Measured by the Daily Inventory of Stressful Events (DISE) will be analyzed using mixed-effects analysis of variance with treatment group by time (Baseline, Post-Intervention, 1-Month Follow-Up) by day in burst (1, 2, …, 7). Models will include fixed (linear) effects for treatment (LARKSPUR vs. control), time (Baseline, Post-Intervention, 1-month Follow-Up), day in burst, and the interactions between treatment group, time, and day in burst. Random intercepts for participants and random slopes for linear day in burst effects will be included to account for within-participant correlation over time. The modified Intention-to-Treat sample described in Section 8.1 will be used. Treatment groups will be compared at Post-Intervention, at 1-Month Follow-Up, and with respect to their change from Post-Intervention to 1-Month Follow-Up using 2-sided tests at the 5% significance level under the null hypothesis of no difference in means/change. All assumptions for the mixed-effects models will be assessed by inspection of plots of within-participant and between-person residuals. Standard errors, degrees of freedom, 95% confidence intervals, and p-values will be adjusted using the Kenward-Roger method to control the significance level [4]. Restricted maximum likelihood (REML) estimation will be used. All analyses are exploratory with the aim of hypothesis generation as described in the Study Protocol given the pilot nature of this study.

11 Safety Analyses

Safety analyses will assess frequencies of All-Cause Mortality, Serious Adverse Events, and Other (Not Including Serious) Adverse Events as required per ClinicalTrial.gov reporting requirements.

Summary statistics will be produced in accordance with Section 9 per treatment group and across all subjects. When calculating the incidence of adverse events, or any sub-classification thereof by treatment, time, severity, etc., each subject will only be counted once, and any repetitions will be ignored; the denominator will be the total population size.

11.1 Extent of Exposure

The summary statistics will be produced in accordance with Section 9.

11.2 Adverse Events

The summary statistics will be produced in accordance with Section 9.

11.3 Deaths, Serious Adverse Events, and other Significant Adverse Events

The summary statistics will be produced in accordance with Section 9.

11.4 Pregnancies

Pregnancies were not assessed as the study population consisted of middle-aged and older adults of at least 50 years of age or older.

11.5 Prior and Concurrent Medications

Baseline pain medication will be assessed with two items: “During the past 30 days, how often have you taken prescription/non-prescription medication for pain?” and “During the past 30 days, how often have you taken non-prescription medication for pain? This may include aspirin (e.g., Bayer), acetaminophen (e.g., Tylenol), ibuprofen (e.g., Advil, Motrin), other NSAIDs (e.g., naproxen a.k.a. Aleve), or something else?” The summary statistics will be produced in accordance with Section 9.

12 Reporting Conventions

P-values ≥ 0.001 will be reported to 3 decimal places; p-values less than 0.001 will be reported as “< 0.001”. The mean, standard deviation, and any other statistics will be reported to two decimal places. Estimated parameters, not on the same scale as raw observations (e.g., regression coefficients) will be reported to two decimal places or two significant figures.

13 References

[1] J.A. Teresi, X. Yu, A.L. Stewart, R.D. Hays, Guidelines for designing and evaluating feasibility pilot studies, Medical Care 60 (2022) 95-103.
[2] A.L. Whitehead, S.A. Julious, C.L. Cooper, M.J. Campbell, Estimating the sample size for a pilot randomised trial to minimise the overall trial sample size for the external pilot and main trial for a continuous outcome variable, Statistical Methods in Medical Research 25 (2016) 1057–1073.
[3] R.A. Parker, C.J. Weir, Non-adjustment for multiple testing in multi-arm trials of distinct treatments: Rationale and justification, Clinical Trials 17 (2020) 562-566.
[4] M.G. Kenward, J.H. Roger, Small sample inference for fixed effects from restricted maximum likelihood, Biometrics 53, no. 3 (1997) 983–97.
[5] R.A. Deyo, K. Ramsey, D.I. Buckley, L. Michaels, A. Kobus, E. Eckstrom, V. Forro, C. Morris, Performance of a patient reported outcomes measurement information system (PROMIS) short form in older adults with chronic musculoskeletal pain, Pain Medicine 17 (2016) 314–324.
[6] D.A. Amtmann, K.F. Cook, M.P. Jensen, W.H. Chen, S.W. Choi, D. Revicki, D. Cella, N. Rothrock, F. Keefe, I. Callahan, J.S. Lai, Development of a PROMIS item bank to measure pain interference, Pain 150 (2010) 173-182.
[7] M. Rose, J.B. Bjorner, J. Becker, J.F. Fries, J.E. Ware, Evaluation of a preliminary physical function item bank supported the expected advantages of the Patient-Reported Outcomes

Measurement Information System (PROMIS), Journal of Clinical Epidemiology 61 (2008) 17-33.
[8] C. Bingham, A. Gutierrez, A. Butanis, V. Bykerk, J. Curtis, A. Leong, A. Lyddiatt, W. Nowell, A. Orbai, S. Bartlett, PROMIS Fatigue short forms are reliable and valid in adults with rheumatoid arthritis, Journal of Patient Reported Outcomes 21 (2019) 1-9.
[9] T. Björgvinsson, S.J. Kertz, J.S. Bigda-Peyton, K.L. McCoy, I.M. Aderka, Psychometric properties of the CES-D-10 in a psychiatric sample, Assessment 20 (2013) 429-436.
[10] S. Cohen, T. Kamarck, R. Mermelstein, A global measure of perceived stress, Journal of Health and Social Behavior 24(4) (1983) 385-396.
[11] B.L. Fredrickson, M.M. Tugade, C.E. Waugh, G.R. Larkin, What good are positive emotions in crisis? A prospective study of resilience and emotions following the terrorist attacks on the United States on September 11th, 2001., Journal of Personality and Social Psychology 84 (2003) 365-376.
[12] D. Watson, L.A. Clark, A. Tellegen, Development and validation of brief measures of positive and negative affect: The PANAS scales, Journal of Personality and Social Psychology 54(6) (1988) 1063-1070.
[13] D.M. Almeida, E. Wethington, R.C. Kessler, The Daily Inventory of Stressful Events: An interview-based approach for measuring daily stressors, Assessment 9 (2002) 41-55.
[14] M.E. Charlson, P. Pompei, K.L. Ales, C.R. MacKenzie, A new method of classifying prognostic comorbidity in longitudinal studies: development and validation, Journal of Chronic Diseases 40(5) (1987) 373–383.
[15] R.D. Boschloo, Raised conditional level of significance for the 2x2-table when testing the equality of two probabilities, Statistica Neerlandica 24 (1970) 1-35.
[16] A. Agresti, Analysis of ordinal categorical data, Wiley (2010) 2nd ed.