Final Statistical Analysis Plan

7 January 2021

Effect of risk-stratified care on disability among adults with low back pain treated in the Military Health System: A randomized clinical trial

NCT03127826

Statistical Analysis Plan

TRIAL FULL TITLE	Validation of the STarT Back Screening Tool in the Primary Care Management of Low Back Pain in the Military Health System: A Randomized Trial of Risk-stratified vs. Usual Care
SAP VERSION	Version 1
SAP VERSION DATE	1/7/2021
TRIAL STATISTICIAN	Emily Poehlein, MB; Cindy Green, PhD
Protocol Version (SAP associated with)	Version 8
TRIAL PRINCIPAL INVESTIGATOR	Dan Rhon, DPT, DSc; Steven George, PhD
SAP AUTHOR(s)	Emily Poehlein, MB; Cindy Green, PhD
ROLES AND RESPONSIBILITIES	Dan Rhon and Steven George: Primary Investigators Emily Poehlein and Cindy Green: Biostatisticians – responsible for finalizing SAP, agreement on best practices and statistical methods for analysis

1 SAP Signatures

I give my approval for the attached SAP entitled *Validation of the STarT Back Screening Tool in the Primary Care Management of Low Back Pain in the Military Health System: A Randomized Trial of Risk-stratified vs. Usual Care* dated 1/7/21

Statistician (Author)
Name: Emily Poehlein, MB

Signature: _____

Date: _____

Statistician Reviewer (As applicable)
Name:

Signature: _____

Date: _____

Principal Investigator
Name:

Signature: _____

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 Error! Bookmark not defined.4

4.2 Scope of the analyses 5

5 Study Objectives and Endpoints 5

5.1 Study Objectives 5

SECONDARY AIMS: 5

5.2 Endpoints 5

6 Study Methods 6

6.1 General Study Design and Plan 6

6.2 Inclusion-Exclusion Criteria and General Study Population 7

Inclusion criteria: 7

Exclusion criteria: 7

6.3 Randomization and Blinding 7

6.4 Study Assessments 97

6.5 Endpoint definitions and variable formats 98

Primary endpoint: Roland Morris Disability Index 98

Pain and Quality of life 98

Biopsychosocial Error! Bookmark not defined.9

Health care utilization 9

Cost effectiveness Error! Bookmark not defined.9

7 Sample Size 109

Sample Size Estimation 109

8 General Analysis Considerations 109

8.1 Timing of Analyses 109

8.2 Analysis Populations 1110

8.2.1 Safety Population 1110

8.2.2 Modified Intention to Treat population 1110

8.3 Covariates and Subgroups 1110

8.3.1 Multi-center Studies 11~~10~~
8.4 Missing Data 11~~10~~
8.5 Multiple Testing 11~~10~~
9 Summary of Study Data 12~~10~~
9.1 Subject Disposition 12~~11~~
9.2 Protocol Deviations 12~~11~~
9.3 Demographic and Baseline Variables 12~~11~~
9.4 Treatment Compliance 12~~11~~
10 Efficacy Analyses 12~~11~~
10.1 Primary Efficacy Analysis 12~~11~~
10.2 Secondary Efficacy Analyses 12
10.2.1 Secondary Analyses of Primary Efficacy Endpoint 12
10.2.2 Exploratory Analyses for Primary Efficacy Endpoint 13~~12~~
10.2.3 Analyses of Secondary Endpoints 13~~12~~
11 Safety Analyses 13~~12~~
11.1 Adverse Events **Error! Bookmark not defined.**~~13~~
11.2 Deaths, Serious Adverse Events and other Significant Adverse Events **Error! Bookmark not defined.**~~13~~
12 Reporting Conventions 13
13 Summary of Changes to the Protocol and/or SAP 13
14 References 15~~14~~
15 Listing of Tables, Listings, and Figures 15~~14~~
15.1 Tables 15~~14~~
Table 1. Summary of protocol deviations 15~~14~~
Table 2. Description of missing primary outcome at 12 months 15~~14~~
Table 3. STarTBack stratification 15~~14~~
Table 4. Baseline characteristics by treatment group 16~~15~~
Table 4. Medical history at baseline by treatment group 16~~15~~
Table 6. CEQ at baseline by treatment group 17~~15~~
Table 7. Summary of Roland Morris at each timepoint (baseline, 6 weeks, 6 months, 12 months) 18~~16~~
Table 5. Pain intensity and satisfaction at baseline by treatment group 18~~16~~
Table 5. Summary of PROMIS at each timepoint (baseline, 6 weeks, 6 months, 1 year) 20~~17~~
Table 6. Summary of NPQ at each timepoint (baseline, 6 weeks) 23~~17~~
Table 7. Summary of PASS at each timepoint (baseline, 6 weeks, 6 months, 1 year) 24~~18~~
Table 7. Summary of SOPA-35 at each timepoint (Baseline, 6 weeks) 24~~18~~

Table 8. Summary of EQ-5D at each timepoint (baseline, 6 weeks, 6 months, 1 year) Error! Bookmark not defined.18

Table 9. Summary of COVID-19 Questionnaire at each timepoint (6 months, 1 year) 2418

Table 10. Summary of OSPRO-YF at all timepoints (baseline, 6 weeks, 6 months, 1 year) 2619

Table 11. Summary of healthcare utilization over 12 months of treatment 2619

Table 12. Summary of cost effectiveness Error! Bookmark not defined.19

15.2 Listings 2619

Listing 1. Adverse events 2619

15.3 Figures 2619

Figure 1. Example CONSORT diagram 2719

3 Abbreviations and Definitions

CEQ	Credibility Expectancy Questionnaire
COVID-19	Coronavirus Pandemic
EQ-5D	EuroQoL
LBP	Low back pain
MHS	Military Health System
MTF	Military treatment facilities
NPQ	Revised Neurophysiology of Pain Questionnaire
OSPRO-YF	Optimal Screening for Prediction of Referral and Outcome Yellow Flags
PASS	Patient Acceptable Symptom Scale
PROMIS	Patient Reported Outcomes Measurement Information Systems
SBST	STarT Back Screening Tool
SOPA-35	Survey of Pain Attitudes-35
QoL	Quality of Life

4 Introduction

4.1 Background/Rationale

Low back pain (LBP) is among the most frequent causes of medical visits and lost-duty time in the Military Health System (MHS). Most clinical practice guidelines recommend only advice and education for all patients with non-specific LBP during the initial weeks of management, with consideration of psychosocial factors and referral to physical therapy recommended only when recovery is delayed. Recent research has demonstrated that the STarT Back Screening Tool (SBST) is useful for classifying patients as being at low, medium, or high risk for experiencing chronic, disabling LBP. Targeted interventions for patients in each risk subgroup have also been developed to address the specific modifiable prognostic indicators identified by the tool. Patients who received stratified care utilizing the SBST demonstrated greater changes in disability, increased quality of life,

and lower healthcare costs compared to patients in the control group at 12 months.

One of the limitations of the STarT Back trial's study design was that there was no standardization of the physical therapy interventions that were delivered. Therefore, it's difficult to ascertain whether the favorable outcomes in the risk stratified group are attributable to superior physical therapy intervention or the overall effectiveness of the stratification process in directing the right patients to physical therapy with or without psychological augmentation. It is also unknown whether a similar stratified care approach will achieve similar results in the primary care management of patients with LBP in the MHS. If these results could be validated, this simple-to-use screening strategy could be implemented practically and efficiently across the MHS, with the expectation that the MHS would realize substantial cost savings and lower disability among MHS beneficiaries with LBP. Therefore, the purpose of this study is to validate the clinical and cost effectiveness of the SBST in the primary care management of patients with LBP in the MHS.

4.2 Scope of the analyses

These analyses will compare efficacy and costs resulting from a risk-stratified approach to a usual care approach in the management of LBP in primary care using the recently developed SBST.

5 Study Objectives and Endpoints

5.1 Study Objectives

PRIMARY AIM:
Specific Aim 1: Compare clinical outcomes of care between risk-stratified care according to the STarT Back Screening Tool and usual care approach in the management of patients with LBP in the primary care setting.

Hypothesis Specific Aim 1: Greater improvements in both short- and long-term clinical outcomes will be observed among patients treated with a risk-stratified versus a usual care approach.

SECONDARY AIMS:
Specific Aim 2: Compare direct costs associated with risk-stratified versus usual care in the management of patients with LBP in the primary care setting.

Hypothesis Aim 2: Direct costs will be lower at 12 months among patients treated with risk-stratified care versus usual care.\

.

5.2 Endpoints

PRIMARY ENDPOINT:
The primary endpoint is self-reported disability (Roland Morris Disability Index) at 12 months.

SECONDARY ENDPOINTS:
Pain and Quality of Life (QoL)

- Patient Reported Outcomes Measurement Information Systems (PROMIS)-57 pain

interference subdomain T Score at 12 months
PROMIS-57 physical function subdomain T Score at 12 months Healthcare Utilization

- Total number of LBP related visits over 12 months of follow-up
- Requiring surgery over 12 months of follow-up
- Requiring another invasive procedure over 12 months of follow-up
- Requiring opioid use over 12 months of follow-up
- Referral to pain clinic over 12 months of follow-up
- Change in SOPA from baseline to 6 weeks
- Change in NPQ from baseline to 6 weeks

6 Study Methods

6.1 General Study Design and Plan

This study is a parallel, two-arm randomized clinical trial with a 1-year follow-up period. The trial will assess superiority of risk-stratified compared to usual care. The study coordinator will remain blind to the patient's treatment group assignment throughout the study. Patients will be randomized to either risk-stratified or usual care, where patients will receive various levels of physical therapy and other management for LBP according to the following table.

Table 1. Components of risk stratified care vs. usual care

Risk Stratified Care	Usual Care
Low, medium and high-risk groups -Evidence-based assessment of LBP according to current clinical practice guidelines, to include limited use of imaging except for patients with "red flags" -Enhanced active management advice emphasizing positive messages about activity, pain relief and work for LBP - Reassurance to address specific concerns related to their LBP and implications on work -Patients given aa copy of the *Back Book* and see a 15-minute video based on the *Back Book* entitled "Get Back Active"	**All patients** -Evidence-based assessment of LBP according to current clinical practice guidelines, to include limited use of imaging except for patients with "red flags" -Active management advice emphasizing positive messages about activity, pain relief and work for LBP -No specific guidance regarding physical therapy referral, thus decision to refer or not will be made by the PCM according to usual practice -Physical therapy according to TBC principles if a referral to physical therapy is made
Low risk group -The above PLUS two-item spinal manipulation screening, with spinal manipulation delivered in primary care if indicated -No referral for ongoing physical therapy	
Medium risk group -Same as low risk group PLUS -Patient is referred for ongoing physical therapy based on the Treatment Based Classification principles for up to 8 visits, 30-60 minute sessions (twice weekly)	
High risk group -Same as low and medium risk group PLUS	

-Patient is referred for ongoing physical therapy using TBC principles for up to 12 visits, 45-60 minute sessions (twice weekly) -Physical therapy is psychologically augmented with the assessment of biopsychosocial risk factors and the adoption of cognitive behavioral principles that explore patient concerns and address unhelpful beliefs and behaviors. These strategies will include tailored education, graded exercise, graded exposure, etc.	

6.2 Inclusion-Exclusion Criteria and General Study Population

Inclusion criteria:

The following inclusion criteria will be used to determine eligibility for the study. Potential candidates must satisfy all criteria and consent to participation to be eligible.

- Males and females who are between the ages of 18-50 years old
- Primary complaint of LBP for any duration, with or without associated radiculopathy
- Can speak and understand English
- Be willing and able to give full, informed written consent

Exclusion criteria:

- Any potentially serious disorders (e.g., cauda equina compression, inflammatory arthritis, compression fracture, malignancy, infection), serious illness, or comorbidity
- Spinal surgery in the past 24 months
- Current pregnancy (or within the last 6 months)
- Already receiving treatment (other than primary care) for this episode of LBP within the last 6 months
- Inability to attend regular treatment sessions
- Pending litigation or a medical evaluation board, ETSing or retiring in the next 10 months

6.3 Randomization and Blinding

After completion of baseline assessments, subjects will be randomized into one of two arms (Group I = Usual Care, Group II= Risk Stratified Care based on the SBST). The method of group assignment will be sequentially numbered opaque sealed envelopes (SNOSE). To minimize the risk of predicting the treatment assignment of the next eligible patient, randomization will be performed in permuted blocks of two or four with random variation of the blocking number. Of the 290 subjects expected for enrollment at SAMMC, ~240 (83%) are expected to be enrolled at BAMC, and ~50 (17%) to be enrolled at WHASC. Randomization envelopes will be managed at the BAMC location, and due to its proximity to WHASC, the research coordinator will be able to provide the randomization number to the research team at WHASC as necessary. Treatment allocation will be tracked and coded locally at each site. Because we are using permuted block randomization by site, there will be a balanced number of patients in each test group for each site. Statisticians responsible for the primary and

secondary analyses will be blinded to the treatment assignment until the SAP is signed and models/tables are programmed.

6.4 Study Assessments

Assessment	Visit / Follow Up (F/U) Interval			
Study Day / period	Same Day	6 weeks	6 months	1 Year
Screening	X			
Informed Consent, discuss Plan, etc.	X			
Randomization	X			
Demographics, History & Physical	X			
Treatment	X	X		
Promis-57	X	X	X	X
OSPRO-YF	X	X	X	X
Credibility Expectancy Questionnaire	X	X	X	X
Medication use & healthcare resource utilization (part of the Data Collection F-U Form)	X	X	X	X
STarTBack Tool	X	X	X	X
Healthcare Utilization (MDR database) for imaging, medication prescription, and specialty referral				X
Quality of Life: *EuroQoL* (EQ-5D)	X	X	X	X
Revised Neurophysiology of Pain Questionnaire (NPQ)	X	X		
Survey of Pain Attitudes (SOPA)	X	X		
Patient Acceptable Symptom Scale	X	X	X	X
Roland Morris Disability Index	X	X	X	X
Coronavirus Pandemic Questionnaire			X	X

6.5 Endpoint definitions and variable formats

Primary endpoint: Roland Morris Disability Index

The Roland Morris Disability Index is calculated from the 24 items from the assessment as the number of times the patient has ticked “today” with greater scores indicating a greater level of disability. It will be analyzed as a count variable.

Pain and Quality of life

PROMIS-57: The PROMIS-57 pain interference and physical function subdomain scores are continuous measures of quality of life with greater scores indicating a lower quality of life. They will be analyzed using sub-component T scores as continuous variables

Health care utilization

LBP-related visits: The number of LBP-related visits including physical therapy, specialist care, etc. will be counted. It will be analyzed as a count variable.

Surgery: The need for LBP-related surgery over 12 months of follow-up will be recorded for each patient. It will be analyzed as a binary variable.

Invasive procedure: the need for an invasive procedure other than surgery such as epidural steroid injection, facet joint injection, trigger point injection, or sacroiliac injection over 12 months of follow-up will be recorded for each patient. It will be analyzed as a binary variable.

Opioid use: the use of opioids at any point over 12 months of follow-up will be recorded for each patient. It will be analyzed as a binary variable. Additional opioid use endpoints include count of unique fills, total days' supply, and morphine milligram equivalents.

Referral to pain clinic: Referral to a pain clinic over the 12 month follow-up period will be recorded for each patient. It will be analyzed as a binary variable.

7 Sample Size

We will enroll 290 consecutive patients seen in primary care with a primary complaint of LBP who consent to participation. Patients will be enrolled in the primary care clinics at four military treatment facilities (MTFs). With a recruitment target of 290 subjects and a conservative recruitment rate of ~30 subjects per month (~10 per site), we will be able to complete recruitment within 1 year. Based on the fact that there are 6-8 new consultations for LBP per day in primary care, and given our previous experience, we do not anticipate having difficulty achieving our sample size within the study timeline.

Sample Size Estimation

The specific aims will be tested by use of pretreatment randomization allocation to low-risk, medium-risk, and high-risk groups in the intervention and control groups. The sample size calculation is based on the ability to detect a between group effect size of 0.3 at the 12-month primary endpoint (Roland-Morris Disability Index). Based on a two-tailed significance level at an alpha level equal to 0.05, 80% power, and allowing for a 25% loss to follow-up, we will aim to recruit 290 participants, 145 in each group overall (Risk Stratified and Usual Care). Numbers below represent estimated recruitment targets at each site. However, one site may over-recruit in the case there are recruitment challenges at another site. Regardless, the total count between both sites will not go over 290.

Estimate Required Sample Size	**234**
Estimate Participant Drop Out / Withdrawal	**56**
Total Enrollment Requirement	**290**

Targeted Enrollment at Each Site	
BAMC	**~240**
WHASC	**~50**

8 General Analysis Considerations

8.1 Timing of Analyses

The final analysis will be performed when the last recruited patient has 12 months of follow-up and after the finalization and approval of this SAP document. The final analysis will be performed on blinded data transferred to statisticians following data cleaning conducted by a study research

coordinator. Data cleaning will include scoring outcomes and checking for missing data/extreme values and outliers.

8.2 Analysis Populations

8.2.1 Safety Population

All subjects who were randomized and participated in at least one post-baseline treatment session will be analyzed in the secondary analysis.

8.2.2 Modified Intention-to-Treat (MITT) population

All subjects who were randomized, participated in at least one post-baseline treatment session, and had baseline and one other Roland-Morris Disability Index measures taken and recorded will be analyzed for the primary aim.

8.3 Covariates and Subgroups

Models for the primary efficacy analysis will include baseline score as a covariate. Secondary models will include active duty status, age, sex, and back pain duration. The CEQ, a 6-item self-report evaluating treatment credibility and expectations for improvement, will be assessed at baseline after treatment group assignment is revealed to provide descriptive information about participants' perceptions of their treatment assignment and optimism for improvement. The Credibility Expectancy Questionnaire (CEQ) may be used as a covariate if perceptions differ meaningfully between groups as participants' initial perceptions of treatment credibility can impact outcomes.

Additionally, for a secondary efficacy analysis, subgroup analyses for, PASS, back pain duration, and active duty status will be conducted using interaction terms.

8.3.1 Multi-center Studies

We will enroll 290 consecutive patients seen in primary care with a primary complaint of LBP who consent to participation. Patients will be enrolled in primary care clinics at Brooke Army Medical Center (BAMC) and Wilford Hall Ambulatory Surgical Center (WHSAC). We will plan to enroll approximately 240 subjects at the BAMC clinic, as this will be the primary site, and plan for approximately 50 subjects at the WHASC clinic. If either of the other sites under-enrolls, then the other site can enroll a higher number, up to the total 290 subjects. We will ensure that total enrollment does not go over 290 between both sites. Since sites are in the same geographical area with similar procedures and patient population, site will not be considered as a covariate in the primary efficacy model but will be evaluated in secondary analyses.

8.4 Missing Data

Missing at random will be judged based on tabulating missing 12 month Roland Morris values by treatment group and demographic characteristics. Mixed effects modelling will be used to allow for missing data. If there is a substantial amount of missing data and missingness is considered to be not random, sensitivity analyses utilizing worst-case imputation will be used for descriptive and inferential assessments to address attrition bias.

8.5 Multiple Testing

Adjustment for multiple testing will be utilized for healthcare utilization endpoints using the Benjamini Hochberg procedure for controlling False Discovery Rate (FDR). The FDR will be controlled at $\alpha = 0.05$.

9 Summary of Study Data

All continuous variables will be summarized using the following descriptive statistics: n (non-missing sample size), mean, standard deviation (SD), median, 25th and 75th percentiles (Q1, Q3), maximum and minimum. The frequency and percentages of observed levels will be reported for all categorical measures using non-missing data. Categorical score data may also be presented as mean (SD). All summary tables will be structured with a column for each treatment in the order (Risk-stratified, Usual) and will be annotated with the total population size relevant to that table/treatment, including any missing observations. SAS version 9.4 will be used for all analyses.

9.1 Subject Disposition

Some withdrawal from the study is expected for various reasons. A CONSORT diagram will be used to describe the number of patients that reached the various stages of the trial and the number of patients that dropped out and for what reasons.

9.2 Protocol Deviations

Protocol deviations are expected to be minimal and will be summarized in tables by study population and treatment.

9.3 Demographic and Baseline Variables

Baseline characteristics recorded at randomization or first treatment administration include demographic characteristics, medical history, current medications, and characterizations regarding LBP. The summary statistics will be produced in accordance with section 9.

9.4 Treatment Compliance

Treatment non-compliance is expected, but difficult to measure. The number of visits will be enumerated for each patient and summarized by treatment risk to determine whether non-compliance was extensive.

10 Efficacy Analyses

10.1 Primary Efficacy Analysis

The primary analysis will compare the Roland Morris Disability Index at 12 months between the two treatment groups. First, Roland Morris Disability Index at each time point will be compared between the two treatment groups in a summary table in the Safety population. Then, using the Modified Intention-to-Treat (MITT) population, treatment effect estimates will be obtained using a repeated measures model including treatment, time (categorical), and an interaction between treatment and time with adjustment for baseline score using an unstructured covariance matrix and degrees of freedom by Kenward-Rodger. Potential distributions for count data include normal, lognormal, negative binomial, and Poisson and will be determined by assessing residuals and other graphical methods. Least-squares means (LS-mean) for each treatment group and the difference between the two will be reported with 95% confidence interval (CI) and two-sided p-value.

10.2 Secondary Efficacy Analyses

10.2.1 Secondary Analyses of Primary Efficacy Endpoint

The analysis of the primary efficacy aim will be repeated with additional adjustment for age, sex, back pain duration, active duty status, and other covariates such as CEQ or PASS that are found to be unevenly distributed between the two treatment groups at baseline.

Additionally, if there is a substantial amount of missing data and missingness is considered to be not random, an additional analysis utilizing worst-case imputation will be implemented.

Lastly, an additional secondary analysis of the primary efficacy endpoint will include further adjustment for therapists' effects using a random effect for primary therapist in the model specified in section 10.1.

10.2.2 Exploratory Analyses for Primary Efficacy Endpoint

For exploratory analyses of the primary efficacy endpoint, subgroup analyses for, PASS, back pain duration, and active duty status will be conducted using interaction terms in the model specified in section 10.1.

10.2.3 Analyses of Secondary Endpoints

PROMIS-57 pain interference and physical function T-scores will be analyzed using the same methods as specified in section 10.1 using a normal likelihood function.

Health care utilization variables will be summarized using tables and compared using non-parametric methods for the number of visits and risk ratios for the other binary outcomes with accompanying p-values.

11 Safety Analyses

Adverse events (non-serious and serious) and deaths are expected to be minimal and will be summarized using listings. If there are a large number of adverse events, they will be tabulated by treatment arm.

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 other than quantiles, will be reported to two decimal places.

13 Summary of Changes to the Protocol and/or SAP

Rationale for Adjustments of Statistical Analysis Plan from Protocol (Version 8, June 30, 2020)

1. Analysis of primary aim

PROTOCOL: "For the primary analyses, imputed datasets will be used for all descriptive and inferential assessments to address attrition bias, generated through multiple imputation (pooled estimates of five imputed datasets) by use of simulation based on a multivariate normal model (numerical variables) and a logistic regression model (categorical outcomes)"

SAP: For the primary analysis, we will plan use a repeated measures model instead of imputation to address missing data since we do not know the extent of missingness in the dataset and imputation

with repeated measures is complicated. In the case that missingness appears to not be random, worst-case imputation will be considered as a secondary sensitivity analysis. Repeated measures models allow all available data to be used and provide less bias than other missing data approaches such as complete case analysis or last observation forward.

14 References

15 Listing of Tables, Listings, and Figures

15.1 Tables

Table 1. Summary of protocol deviations

Deviation, N (%)[a]	Risk stratified	Usual

[a] Percent of all deviations

Table 2. Number of visits by risk stratification level

	Low	Medium	High
Number of visits			
N			
Mean (SD)			
Median (Q1-Q3)			

Table 3. Description of missing Roland Morris Index at 12 months

	Missing (n =)	Not missing (n =)
Age, Mean (SD)		
Sex, N(%)		
Female		
Male		
MtF Female		
FtM Male		
Race, N(%)		
American Indian/Alaska Native		
Asian		
Black or African American		
Native Hawaiian or Pacific Islander		
White		
More than one race		
Other		
Treatment, N(%)		
Risk stratified		
Usual		

Table 4. STarTBack stratification

	Low Risk	Medium Risk	High Risk
Total score			

N			
Mean (SD)			
Median (Q1-Q3)			
Min - Max			
Sub score Q5-9			
N			
Mean (SD)			
Median (Q1-Q3)			
Min - Max			

Table 5. Baseline characteristics by treatment group

	Risk-stratified (n =)	Usual (n =)
Age		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		
Sex, N(%)		
Female		
Male		
MtF Female		
FtM Male		
Race, N(%)		
American Indian/Alaska Native		
Asian		
Black or African American		
Native Hawaiian or Pacific Islander		
White		
More than one race		
Other		
Back pain duration		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		
Rank, N(%)		
Education level, N(%)		
Marital status, N(%)		
Employment status, N(%)		
Tobacco use, N(%)		
Activity, N(%)		

Table 6. Medical history at baseline by treatment group

	Risk-stratified (n =)	Usual (n =)
Musculoskeletal injury, N(%)		
Type of injury, N(%)		

Other medical issues, N(%)		
Current medications, N(%)		

Table 7. CEQ at baseline by treatment group

	Risk-stratified (n =)	Usual (n =)
How logical does the therapy offered to you seem?		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		
How successful do you think this treatment will be in reducing your limitations due to back pain?		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		
How confident would you be in recommending this treatment to a friend who experiences similar problems?		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		
By the end of the therapy period, how much improvement in your limitations due to back pain do you think will occur?		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		
At this point, how much do you really feel that the therapy will help to reduce your limitations due to back pain?		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		
By the end of the therapy period, how much improvement in your limitations due to back pain do you really feel will occur?		

N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		

Table 8. Summary of Roland Morris at each timepoint (baseline, 6 weeks, 6 months, 12 months)

	Risk-stratified (n =)	Usual (n =)
Summary score		
Baseline		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		
6 weeks		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		
6 months		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		
12 months		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		

Table 9. Roland Morris Score at 12 months

	Risk-stratified	Usual	Difference	p-value
Least-Squares Mean (95% CI)				

Table 10. Pain intensity and satisfaction at baseline by treatment group

	Risk-stratified (n =)	Usual (n =)
Current pain,		
Baseline		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		
6 week		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		

6 month		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		
12 month		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		
Pain at worst, Mean (SD)		
Baseline		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		
6 week		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		
6 month		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		
12 month		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		
Pain at best, Mean (SD)		
Baseline		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		
6 week		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		
6 month		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		
12 month		

N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		

Table 11. Summary of PROMIS-57 Pain Intensity at each timepoint (baseline, 6 weeks, 6 months, 1 year)

	Risk-stratified (n =)	Usual (n =)
Summary score		
Baseline		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		
6 weeks		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		
6 months		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		
12 months		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		

Table 12. PROMIS-57 Pain Intensity Score at 12 months

	Risk-stratified	Usual	Difference	p-value
Least-Squares Mean (95% CI)				

Table 13. Summary of PROMIS-57 Physical Function at each timepoint (baseline, 6 weeks, 6 months, 1 year)

	Risk-stratified (n =)	Usual (n =)
Summary score		
Baseline		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		
6 weeks		
N		
Mean (SD)		
Median (Q1-Q3)		

Min - Max		
6 months		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		
12 months		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		

Table 14. PROMIS-57 Physical Function Score at 12 months

	Risk-stratified	Usual	Difference	p-value
Least-Squares Mean (95% CI)				

Table 15. Summary of PROMIS-57 Sleep Disturbance at each timepoint (baseline, 6 weeks, 6 months, 1 year)

	Risk-stratified (n =)	Usual (n =)
Summary score		
Baseline		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		
6 weeks		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		
6 months		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		
12 months		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		

Table 16. Summary of PROMIS-57 Fatigue at each timepoint (baseline, 6 weeks, 6 months, 1 year)

	Risk-stratified (n =)	Usual (n =)
Summary score		
Baseline		
N		
Mean (SD)		

Median (Q1-Q3)		
Min - Max		
6 weeks		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		
6 months		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		
12 months		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		

Table 17. Summary of PROMIS-57 Social Participation at each timepoint (baseline, 6 weeks, 6 months, 1 year)

	Risk-stratified (n =)	Usual (n =)
Summary score		
Baseline		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		
6 weeks		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		
6 months		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		
12 months		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		

Table 18. Summary of PROMIS-57 Depression at each timepoint (baseline, 6 weeks, 6 months, 1 year)

	Risk-stratified (n =)	Usual (n =)
Summary score		
Baseline		
N		

Mean (SD)		
Median (Q1-Q3)		
Min - Max		
6 weeks		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		
6 months		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		
12 months		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		

Table 19. Summary of PROMIS-57 Anxiety at each timepoint (baseline, 6 weeks, 6 months, 1 year)

	Risk-stratified (n =)	Usual (n =)
Summary score		
Baseline		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		
6 weeks		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		
6 months		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		
12 months		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		

Table 20. Summary of NPQ at each timepoint (baseline, 6 weeks)

	Risk-stratified (n =)	Usual (n =)
Summary score		
Baseline		
N		

Mean (SD)		
Median (Q1-Q3)		
Min - Max		
6 weeks		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		

Table 21. Summary of PASS at each timepoint (baseline, 6 weeks, 6 months, 1 year)

	Risk-stratified (n =)	Usual (n =)
Do you consider the current state of your low back satisfactory, you are ok if nothing changes? N(%)		
Baseline		
6 week		
6 month		
12 month		

Table 22. Summary of SOPA-35 at each timepoint (Baseline, 6 weeks)

	Risk-stratified (n =)	Usual (n =)
Summary score		
Baseline		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		
6 weeks		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		

Table 23. Summary of COVID-19 Questionnaire at each timepoint (6 months, 1 year)

	Risk-stratified (n =)		Usual (n =)	
	6 months (n =)	12 months (n =)	6 months (n =)	12 months (n =)
Ability to get health care for back pain, N(%)				
Reduced my ability to get health care a lot				
Reduced my ability to get health care a little				
Not affected my ability to get healthcare				
Improved my ability to get healthcare				

Social support, N(%)				
A lot less social support				
A little less social support				
My social support was not affected				
Improved my social support				
Mental and emotional health, N(%)				
My mental health is a lot worse				
My mental health is a little worse				
Not affected my mental health				
Improved my mental health				
Level of back pain, N(%)				
My back pain is a lot worse				
My back pain is a little worse				
Not affected my back pain				
Improved my back pain				
Ability to continue treatment goals using self-management or other treatment strategies, N(%)				
I have made no progress on treatment goals				
I have made little progress on treatment goals				
I have made some progress on treatment goals				
I have made significant progress on my treatment goals				
Which of the following statements best describes your personal experience with the coronavirus?, N(%)				
I think I was sick with coronavirus but did not go to the hospital				
I don't think was sick with the coronavirus				
Unsure				
I was sick with the coronavirus and either stayed overnight or was admitted to the hospital				
Which of the following statements best describes the experience of people you live with?, N(%)				
I think someone I live with was sick with the coronavirus but did not go to the hospital				
I don't think anyone I live with was sick with the coronavirus				
Unsure				

Someone I live with had an overnight stay or was hospitalized while they were sick				

Table 24. Summary of OSPRO-YF at all timepoints (baseline, 6 weeks, 6 months, 1 year)

	Risk-stratified (n =)	Usual (n =)
Summary score		
Baseline		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		
6 weeks		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		
6 months		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		
6 months		
N		
Mean (SD)		
Median (Q1-Q3)		
Min - Max		

Table 25. Summary of healthcare utilization over 12 months of treatment

	Risk-stratified (n =)	Usual (n =)
Number of visits, Mean (SD)		
Number of visits (categorical), N(%)		
Required surgery, N(%)		
Required an invasive procedure, N(%)		
Required opioids, N(%)		
Referred to pain clinic, N(%)		

15.2 Listings

Listing 1. Non-serious Adverse events

Listing 1. Serious Adverse events

Listing 1. Deaths

15.3 Figures

Figure 1. CONSORT diagram
Figure 2. Roland Morris Score over time by treatment group
Figure 3. PROMIS-57 Pain Intensity over time by treatment group
Figure 4. PROMIS-57 Physical Function over time by treatment group
Figure 5. SOPA over time by treatment group
Figure 6. NPQ over time by treatment group