Proposal for A Trial to Test the Efficacy and Safety of Sprifermin Compared to Physical Therapy in Knee Osteoarthrosis

An active-controlled multi-centered randomized double-blinded phase III study

April 14, 2021
P8140: Introduction of Randomized Clinical Trials
Xiangyi Liu – xl3048

TABLE OF CONTENTS

1 Introduction and Background

Knee osteoarthrosis (OA) is an increasing prevalent joint disorder common in middle-age and older adults, involving symptoms such as knee pain and stiffness[1]. As a disease utilizes a substantial healthcare resources, it affects ~10% men and ~18% women ages ≥60 years old worldwide[2]. It occurs when the cartilage that cushioning the bones gradually degenerates. Besides, knee OA causes deterioration of the connective tissues that hold the joint together and connect the muscle to the bone[3].

Treatment for knee OA can be broadly classified into three categories: analgesic medications, physical therapy, surgical and other procedures, such as intra-articular injections to provide cushioning within the joint and to reduce the symptoms[4].

Both corticosteroid injection (CSI) and physical therapy are commonly subscribed as the standard of care for knee OA. However, there are conflicting reports the limited period of benefit of CSI and possible complications such as joint infection and accelerated degradation of articular cartilage of an invasive procedure[5]. Physical therapy relieves pain by strengthening the muscle around the joint and increasing the flexibility, which may lead to more durable relive and greater functional improvement[6].

Sprifermin is a newly developed medication that has been proved to contribute to the repair and generation of cartilage by increasing the production and development of bone cells in the Phase II trial, which demonstrated the potential of long-term improvement for individuals with knee OA[7]. Considering the different focuses of Sprifermin and physical therapy, one on the regeneration of cartilage, one on rehabilitation of joint muscle, which are two main aspects of knee OA treatment, we become interested in the efficacy of the regimen of both approaches on treating the disease.

The trial will be a multicentered, randomized, double-blinded, active-controlled, parallel-group Phase III trial. The target population is U.S. adults of either sex of any race between 40 and 85 years old. Patients need to meet the American College of Rheumatology (ACR) clinical criteria for a knee OA diagnosis.

The clinical question of the trial is whether the regimen of Sprifermin plus physical therapy is superior to physical therapy alone in reducing the mean score from baseline in the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) after one-year individual follow-up period. Higher the score indicates worse function, pain, and stiffness.

100 mcg active Sprifermin will be administrated to patients of the intervention group as intra-articular injections once a week for consecutive three weeks for two cycles, that is at week 0, 1, 2 in Cycle 1; at week 26, 27, 28 in Cycle 2 at an interval of 6 months[8]. Patients in the control group will receive placebo Sprifermin identical in appearance at the same frequency as intra-articular injections.

Patients in both groups will receive physical therapies tailored according to their diagnosis of knee OA and the functional examinations at the screening. They will undergo a total of twelve 40-minute physical therapy sessions: 8 of them will be conducted in the first 8 weeks, 2 sessions at week 12, 13, 14, 1 session at week 26, 1 session at week 38, 39, 40[11].The manual physical therapy protocol consists of procedures that intend to reduce the stiffness of the soft tissue structures and increase the quality and range of motion of the affected knee. Additional individualized physical therapy will be led by physical therapists based on the examination and interview with the patients at the first session.

2 Objectives

i) Primary Outcome

The change of WOMAC score from baseline at Year one [Time Frame: Baseline, Year 1 (Week 52)]

- *Null Hypothesis*:

- There is no difference in changes of WOMAC score at Year one between Sprifermin plus physical therapy regimen and physical therapy alone at Year 1 at 0.05 significance level.

- *Alternative Hypotheses*:

- Sprifermin plus orthopedic manual physical therapy is superior in reducing WOMAC score at Year one to physical therapy alone at Year 1 at 0.025 significance level.
- Orthopedic manual physical therapy alone is superior in reducing WOMAC score at year one to Sprifermin plus physical therapy at Year 1 at 0.025 significance level.

- *Clinical Interest:*

The clinical interest is whether the regimen of Sprifermin plus physical therapy is superior to physical therapy alone in treating knee OA at Year 1.

ii) Secondary Outcomes

- **Change in the scores of 36-Item Short Form Survey (SF-36) questionnaire from baseline at Year one. [Time Frame: Baseline, Year 1 (week 52)]**

  Null Hypothesis: There is no difference in the mean change in the scores of SF-36 between the regimen of Sprifermin and physical therapy group and the physical therapy alone group at Year 1 at 0.05 significance level.

Alternative Hypotheses:

- Sprifermin plus physical therapy regimen is superior in increasing the mean SF-36 score than physical therapy alone at Year one at a significance level of 0.025.
- Physical therapy alone is superior in increasing the mean SF-36 score than the regimen of Sprifermin and physical therapy at Year one at a significance level of 0.025.

- **Change from baseline in cartilage thickness in the total femorotibial joint at Year one [Time Frame: Baseline, Week 12, 26, 52]**

  Null Hypothesis: There is no difference in the mean cartilage thickness in the total femorotibial joint between the regimen of Sprifermin and physical therapy group and the physical therapy alone group at Year 1 at 0.05 significance level.

  Alternative Hypotheses:

- Sprifermin plus physical therapy regimen is superior in increasing the mean cartilage thickness in the total femorotibial joint than physical therapy alone at Year 1 at a significance level at 0.025.
- Physical therapy alone is superior in increasing the mean cartilage thickness in the total femorotibial joint than the regimen of Sprifermin and physical therapy at Year 1 at a significance level of 0.025.

- **Change from baseline in the 40-meter fast-paced walk test (40mFPWT) at Week 12, 26, 38, 52. [Time Frame: Baseline, Week 12, 26, 38, 52]**

Null Hypothesis: The mean time used for a 40mFPWT does not differ in the regimen of Sprifermin and physical therapy group and the physical therapy alone group at Year 1 at a significance level of 0.025.

Alternative Hypotheses:

- Sprifermin plus physical therapy regimen is superior in decreasing the mean time used for 40mFPWT than physical therapy alone at Year 1 at a significance level of 0.025.
- Physical therapy alone is superior in decreasing the mean time used for 40mFPWT walk test than the Sprifermin and physical therapy regimen at Year 1 at a significance level of 0.025.

- **Change from baseline in the 6-minute walk test (6mWT) at Week 12, 26, 38, 52. [Time Frame: Baseline, Week 12, 26, 38, 52]**

The 6-minute walk test is a functional task that asks patients to walk as far as possible. The distance (in centimeters) covered is recorded as performance assessments.

Null Hypothesis: The mean distance covered in the 6mWT does not differ in the regimen of Sprifermin and physical therapy group and the physical therapy alone group at Year 1 at a significance level of 0.025.

Alternative Hypotheses:

- Sprifermin plus physical therapy regimen is superior in decreasing the mean time used for 6mWT than physical therapy alone at Year 1 at a significance level of 0.025.

- Physical therapy alone is superior in decreasing the mean time used for 6mWT than the Sprifermin and physical therapy regimen at Year 1 at a significance level of 0.025.

iii) Safety Outcomes

Safety outcomes should be monitored continuously. All SAEs and AEs should be recorded by Medical Dictionary for Regulatory Activities (MEDRA). Key safety events are reported by patients immediately they occur. Below listed safety outcomes are all proportional variable and are considered as clinical-relevant and representative to the trial.

- **Infections and infestations**
- Number of patients experienced sepsis arthritis
- **Musculoskeletal and connective tissue disorders**
- Number of patients experienced decreased range of motion in knees
- Number of patients experienced knee pain
- Number of patients experienced knee swelling

3 Trial Design

i) RCT Features

Prospective, Intervention & Control Group - The proposed trial is a multicentered, double-blinded, active-controlled, parallel-grouped Phase III trial. A placebo Sprifermin plus physical therapy as the standard of care for knee OA is utilized as an active control. The trial is prospective in nature, and it assess the association between the regimen of interest, which is active Sprifermin plus physical therapy, and the efficacy on functional improvement for individuals with knee OA.

Randomization & Double Blinding - Patients are randomized to the regimen of Sprifermin plus physical therapy or physical therapy alone in a 1:1 ratio on the basis of a stratified (strata defined by study site) permuted block-randomization. Patients and investigators will not be informed of the treatment allocation during the trial unless there is a special need, such as a medical emergency, to reveal the blinding. The outcome will be assessed by on-site investigators who are not involved in care of the patients. The on-site assessors will not be allowed to witness any patients-care provider interactions to avoid potential bias in the assessment as much as possible.

Intent-To-Treat (ITT) Analysis & Complete Follow Up – Patients will be informed that manual physical therapy is a well-established treatment for knee OA, and it is proved to provide short-term benefits in reducing pain and improving physical performance[9]. Long-term consecutive treatment sessions will be required to main the result. Sprifermin is shown to not only increase the cartilage thickness, but also reduces cartilage loss in phase II trial[10]. Patients will be encouraged by sponsors to participate and stay for the whole course of the trial to achieve less than 1% loss to follow-up. Post hoc primary analysis will be conducted based on the treatment assignments.

ii) Blinding

Patients, healthcare providers, data collectors and on-site outcome assessors will be blinded from treatment allocation; the Data and Safety Monitoring Board (DSMB) and data analysts will be informed of the treatment allocation.

A randomization list for each study site will be generated before patient enrollment. As each patient is randomized to the trial, he/she will be allocated to either the intervention group (active Sprifermin and physical therapy) or the control group (placebo Sprifermin and physical therapy).

Each patient will be assigned a unique ID, which identifies the study site and the patient. The ID will remain the same if the patient switch to another study site during the trial. The treatment assignments will be recorded as Group A (intervention group) and Group B (control group) as a record for healthcare providers for the purpose of future treatment administration. Before randomization begins, two sets of pre-filled 100 mcg syringes labeled as 'Group A' and 'Group B' will arrive at study sites. Placebo Sprifermin is formulated to be identical in appearance and package to the active Sprifermin so that neither patients nor the healthcare providers can distinguish between them. On-site outcome assessors will not participate in any treatment administration or observe any interaction between patients and healthcare providers to minimize bias in the assessment. Data collected will be blinded with regard to acquisition order and treatment assignments.

The potential of blinding will be exanimated before the trial. Two volunteers who will not participate in the trial will be asked to distinguish between the active and placebo Sprifermin independently.

An end-of-trial test of blindness will be conducted after the treatment administration, which will minimally affect an individual's behaviors during the trial. Patients will be asked to guess their treatment assignments, and their answers will be selected from 'active/placebo/do not know'. If the blinding is successful, the percentage of answering 'do not know' should be particularly high, or the proportions of correct answers and incorrect answers are approximately the same[12].

iii) Randomization

In this trial, patients will be randomized according to a permuted block randomization scheme with block sizes of 4, 6, and 8, stratified by study sites. With the scheme, the randomization

lists of each study site can be generated before patient enrollment. The proportions of block sizes are prespecified as following: 60% for block size 4, 30% for block size 6, and 10% for block size 8. Random treatment assignments are generated by randomly selecting half of the assignments to be to the intervention group (active Sprifermin plus physical therapy regimen) and then allowing the remaining half of assignments to be the control group (placebo Sprifermin plus physical therapy regimen). As each patient is randomized to the trial, the patient receives the next sequential assignment on the randomization list of his/her specific site. The use of random block size prevents the allocation of treatment from being speculating. Stratification by study sites brings two advantages to the design. Firstly, it ensures that poor enrollment or even discontinuation of participating of one site will not affect the overall balance of treatment groups. Secondly, since the trial is multicentered, stratification by sites prevents the imbalance of treatment responsiveness caused by the difference in risk factors at different sites.

iv) Inclusion and Exclusion Criteria

Main Inclusion Criteria –

- Age from 40 to 85 years; of either sex (knee OA prevalent population)
- Nationality: United States. All races included.
- Meet the American College of Rheumatology (ACR) clinical criteria for a diagnosis of knee OA
- Radiographic stage 2 or stage 3 disease in the target knee as determined by Kellgren/Lawrence (K/L) scale scores from weight-bearing fixed flexion radiographs †
- A history of pain due to Osteoarthritis in the target knee for at least 6 months

- Need for regular analgesic medication(s) for knee pain on more than half of the days in the previous months †
- WOMAC scores of 24–72 (mild to severe, but not extreme) †
- Written consent form

† *Ensure the difference at baseline is not significant enough to impact the primary outcome.*

Main Exclusion Criteria –

- Clinical signs of inflammation (redness) in the target knee *
- Planned knee surgery (affecting either the target or the contralateral knee) within the next two years
- Have complaints of low back pain or other lower extremity joint pain that affects function at the time of recruitment *
- Have a history of neurological disorders that would affect lower extremity function
- Participation in another clinical trial within the 30 days before screening *
- Pregnant women

**Not able to distinguish whether the clinical symptoms are caused by the intervention of interest*

v) Enrolling Centers

Centers enrolling the trial include hospital clinics in Northwell Health, which is the largest academic health system in New York.

vi) Data Coordination and Trial Management

Data Coordination Center (DCC) plays a pivotal in the successful execution of clinical trials. An effective DCC collaborates closely with clinical investigators on protocol development, statistical design strategy, data collection, verification and storage, manuscript preparation, and coordination with DSMB, Protocol Review Committee (PRC), and other committees for quality assurance. The proposed trial as a multicenter study will require a DCC that is responsive of all phases of the coordinated research study.

Clinical Trial Management (CTM) resources are systems that are employed to manage clinical trials in terms of financial planning, trial progress planning, electronic data capture, adverse events reporting, patient information organizing. There are many choices of CTM software systems online, and for the proposed trial, Flex Databases can be implemented for trial management.

vii) Sidedness of Test

The hypothesis test on the primary outcome, which is the WOMAC score at year one, will be conducted at the alpha = 0.05 two-tailed level of significance. The null hypothesis is there is no difference in the mean WOMAC score change from baseline between the intervention and control groups. Since the superiority of the intervention and control groups is to be determined, there are two alternative hypotheses, with each one of them tested at the alpha = 0.025 one-tailed level of significance. The first alternative hypothesis is, the intervention (Sprifermin plus physical therapy) is superior in reducing WOMAC score at year one; the second alternative hypothesis is, the control (physical therapy) is superior in reducing WOMAC score at Year 1.

4 Data Collection and Patient Follow-up

i) Outcome Details

Note: If the patient is scheduled to receive an intra-articular injection in the same week as the test, the injection should occur after tests to obtain as much objective result.

Primary Outcome*:*

- *Change in the scores of WOMAC from baseline at Year one. [Time Frame: Baseline, Year one (Week 52)]*

The primary outcome is a continuous variable, and the instrument used is the WOMAC questionnaire. It is initially developed for patients with knee and/or hip OA, and it is widely used as a measure to evaluate clinically important, patient-relevant changes in health status as a result of treatment intervention. As a validated instrument, its reliability (reproducibility), validity (correlation with other instruments that measure similar constructs), and responsiveness (the ability to detect the change when it truly occurred) among patients with knee OA have been validated with substantial evidence[13].

The WOMAC questionnaire consists of a total of 24 questions and evaluates patient's health status from three measures: pain (5 questions), articular stiffness (2 questions), and physical function (17 questions). The sum of the score for each question will be transformed into 0-100 scale, and the higher the score means worse the condition. A negative value in change indicates performance improvement[14].

Patients will be asked to answer the questionnaire twice, during the screening period and after the last treatment administration in Week 52. Patients enrolled in the trial will use the WOMAC score obtained during the screening period as the baseline score.

When patients arrive at the study site, they will receive the questionnaire from clinical investigators along with other health assessments. It will take 5 to 10 minutes to complete, and the Data Coordinator will collect the form with the patient's self-reported scores with his/her name and signature on it. Patient information and their answers to every question will need to be recorded. Although the primary outcome is the total score change from baseline, data on three measures may provide investigators with information on which aspect contributes the most to the overall improvement.

Secondary Outcomes*:*

- *Change in the mean score of SF-36 questionnaire from baseline at Year one [Time Frame: Baseline, Year one (Week 52)]*

  It is a continuous variable, and the instrument used is the SF-36 questionnaire.

SF-36 is a widely used, well-studied, self-reported measure of health status. It comprises 336 questions which cover eight domains of health: 1) Bodily pain; 2) Mental Health; 3) Physical role; 4) Physical functioning; 5) Social functioning; 6) Health transition; 7) Emotional health; 8) General health. The higher score of SF-36 indicates a more desirable health status. The internal consistency, reproducibility, validity and discriminatory power of SF-36 have been tested and confirmed with 1582 participants[15].

Patients will be asked to answer the questionnaire twice when they come to the study site, during the screening period and after the last treatment administration in Week 52. The questionnaire will take about 5 minutes to complete (tick boxes). Using the predefined scoring key, investigators will transform patients' responses to a numeric scale as scores of SF-36. Data Coordinators should collect both the patients' raw answers and scores, along with the patients' IDs.

The scores will be used for statistical analysis, and the raw answers should be carefully stored for future references.

- *Change from baseline in cartilage thickness in the total femorotibial joint at baseline, week 12, 26, 52 [Time Frame: Baseline, Week 12, 26, 52)]*

It is a continuous variable and will be evaluated by quantitative magnetic resonance imaging (qMRI).

16 subregions in cartilage will be segmented, and the summary scores of both the positive and negative changes in subregions will be summarized as the total cartilage thinning sum score and the total cartilage thickening sum scores. In the way, changes in either direction can be captured[10]. Unlike the conventional measure on the average cartilage change, the instrument employed takes into account the different direction of change in the cartilage thickness across different patients, which cancel each other out when the mean change across patients is reported. The reproducibility and validity of the instrument has been published with substantial evidence[16].

Four measurements will be taken at the baseline, Week 12, 26, and the end of study (Week 52). Clinical investigators at the study site will acquire the coronal magnetic resonance images of patients. Segmentation will be performed in the order different from acquisition by on-site assessors who are blind to treatment allocation. Data Coordinators will collect and record the patients' IDs, segmentations and the corresponding summary scores obtained at four time points, but only the change between baseline and Week 52 will be used in the hypothesis testing to limit multiple comparison and because a 12-month interval was proved to represent a robust and persistent treatment effect in clinical trials[17].

Note: Two outcomes listed below are performance-based tests. Proper time interval (at least one hour) should be allowed before the first and the second test.

- *Change from baseline in the 40-meter (4×10m) fast-paced walk test at Week 12, 26, 38, 52. [Time Frame: Baseline, Week 12, 26, 38, 52]*

It is a continuous variable. A stopwatch for timing and a 10 m marked with space to turn safely at each end are required.

The 40mFPWT is a functional task that measures the time (in seconds) spent for the patient to walk 40 meters as fast as possible, without running. It is an objective, performance-based test that measures the physical ability in patients with knee OA. Its test-retest reliability and measurement error at an acceptable level have been previously published[18].

The test at the study site will be conducted 5 times through the trial at baseline, Week 12, 26, 38, and 52 (end of study). Investigators need to prepare a 10-meter walkway marked with bright colored tape at each end. Place a cone approximately 2 meters before the start mark and 2 meters beyond the finish mark for safe turning. Investigators will inform patients of wearing comfortable, flat walking footwears before they arrive at the site. If there are multiple patients ready for the test simultaneously, a maximum of four patients are allowed to take the test together for safety concerns. Before the actual test begins, patients can take a practice trial for 1-2 turns to ensure a correct understanding. Patients will be asked to walk as quickly but as safely as possible along the marked walkway and turn around the cone three times for a total distance of 40 meters. Regular walking aid will be allowed and recorded.

Data Coordinators will collect data on the time used for 40mFPWT, week in which the test is conducted, walking aids if applicable, and the patients' IDs.

- *Change from baseline in the 6-minute walk test (6mWT) at Week 12, 26, 38, 52. [Time Frame: Baseline, Week 12, 26, 38, 52]*

It is a continuous variable. A flat walking area, preferable >20 m in length, and a stopwatch for timing are required.

The 6mWT is a performance-based test of patient's aerobic walking ability over longer distances. The maximal distance cover in the 6-minutes period is recorded. It is a commonly used method to assess the physical function of patients with knee OA, and its test-retest reliability, measurement error level and the correlation with disease severity have been validated[18,19].

A flat walking area (hallway) without traffic, with distance interval marked every 3 meters, should be prepared. Place a cone to indicate the turn points. Chairs for resting can be placed at the turning point if needed. Investigators will inform patients of wearing comfortable, flat walking footwears before they arrive at the site. If there are multiple patients ready for the test simultaneously, a maximum of four patients are allowed to take the test together for safety concerns. The timer will not be stopped if the patient takes a rest.

Data Coordinators will collect data on patients' IDs, walking aids if applicable, distance covered in 6mWT, and the week of measurement.

ii) Data Collection Mechanism

A web-based data management system with Electronic Case Report Forms (eCRFs) will be implemented. Questionnaires filled by hand will require investigators to enter patients' answers into the system.

iii) Schedule of Visits

Table 1. Schedule of Visits

Milestones	Screening	Baseline			
Visit Number	V1 to V2	V3	V4	V5	V6 to V10
Week	Wk-2 to Wk-1	Wk0	Wk1	Wk2	Wk3 to Wk7
Day	d-14 to d-1	d1	d8	d15	d22 to d56
Procedures&eCRFs:					
Trial informed consent	X				
Inclusion & exclusion	X				
Medical History	X				
Medication History	X				
Demographics	X				
SF-36 questionnaire		X			
WOMAC questionnaire [a]	X				
40mFPWT	X				
6mWT	X				
Physical therapy planning	X				
Cartilage qMRI	X				
Randomization		X			
Treatment:					
Administration training	X				
Study drug dispensing	X				
Study drug administer		X	X	X	
Physical therapy		O	O	O	X [b]
Safety Assessment:					
Body weight	X				
Blood pressure and pulse	X				
Physical exam	X				
Collect adverse event		X	X	X	X

(Table continues)

Milestones			Last Treatment Dispensing	EOS, EOT
Visit Number	V11 to V12	V13	V14	V15
Week	Wk12 to Wk14	Wk26	Wk38 to Wk40	Wk52
Day	d85 to d98	d183	d267 to d280	d365
Procedures&eCRFs:				
SF-36 questionnaire				X
WOMAC questionnaire [a]				X
40mFPWT	X	X	X	X
6mWT	X	X	X	X
Cartilage qMRI	X	X		X
Treatment:				
Study drug administer		X		
Physical therapy	X [c]	X	X	
Safety Assessment:				
Body weight				
Blood pressure and pulse				
Physical exam				
Collect adverse event	X	X	X	X

X = Required

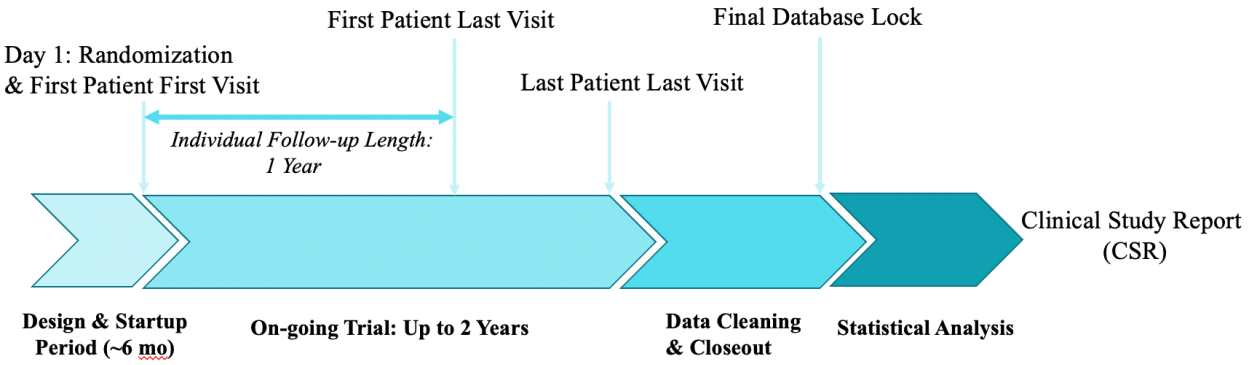
O = Optional

[a] WOMAC scores acquired at the screening period will be used as baseline scores for eligible patients.

[b] A total of 8 physical therapy sessions are required from Week 0 to Week 7. Thereby during the period, the arrangement of physical therapy sessions is not restricted to a particular week.

[c] Two physical therapy sessions are required from Week 12 to Week 14. Thereby during the period, the arrangement of physical therapy sessions is not restricted to a particular week.

iv) Trial Timeline



5 Statistical Considerations

i) Type of Outcomes

The primary outcome (mean WOMAC score change from baseline) is a continuous variable. The hypothesis testing will be performed using two-sample independent t test for means assuming equal variances. The hypothesis testing will be two-sided. The null hypothesis is that there is no difference in the primary outcome between the two groups. The alternative hypotheses have two components, and each tests the superiority of on to the other for the intervention and control groups. The experimental-wise type I error is 0.05.

ii) Power Calculation

From a previously published study on the efficacy of physical therapy that is similar in the schedule and content to the proposed trial on individuals with knee OA, the mean total WOMAC score is 88.1 (12.8) at baseline and the mean change of WOMAC score at Year one is -33.9 (SD=11.8), which is an approximately 38.5% improvement from baseline[20]. Since there is no previous trial on the regimen of Sprifermin plus physical therapy comparing to placebo, the

proposed trial approximates the addition of Sprifermin to physical therapy treatment will lead to an extra 20% reduction in the mean WOMAC change at Year one, which is -40.68.

The unadjusted effect size is 6.78. After assuming a 5% crossover rate and 15% noncompliance rate for both groups, the adjusted effect size is 4.785 (Appendix Sect.1).

Because the mean WOMAC change decreases for the intervention group and increases for the control group after taking into account crossover and noncompliance, the effect size increases after adjustment. The sample size will be underestimated if the unadjusted effect size is used for calculation.

iii) Sample Size

Sample size calculation is performed using Power Analysis and Sample Size 2021 software (PASS Version 21.0.2). A total sample size of 318 (159 per group) is needed to achieve 80% power for an effect size of 4.785 unit increase in the mean WOMAC change at Year one, at a significance level of 0.05 two-sided. The standard deviation is set to be 15, rather than 11.8, which is the standard deviation corresponding to the mean WOMAC change from the published study, to obtain a more conservative sample size.

Data involved in the calculation are: alpha as 0.05; power as 0.80; mean for the control group as 35.0475; mean for the intervention group as 39.8325; standard deviations for two groups as 15; the number of interim looks as 5; fractions of information obtained at each interim look as 0.2, 0.4, 0.6, 0.8, and 1.0.

Group-Sequential Tests for Two Means (Legacy)

Numeric Results for Two-Sided Hypothesis Test of Means

Target Power	Actual Power	N1	N2	N	Mean1	Mean2	S1	S2	Alpha
0.8	0.80228	159	159	318	39.8325	35.0475	15	15	0.05

iv) Sensitivity Analysis

The sample size estimation for the sensitivity analysis is conducted by applying different standard deviations (s) and effect size (d). Holding the effect size fixed, a change in the standard deviation of size 0.5 will lead to a sample size change of approximately 20. Keeping the standard deviation constant, a change in the effect size of 0.215 units will lead to a sample size change of roughly 25 to 35. The sensitivity analysis provides us with information on the sensitivity of sample size and the amount of change will be expected if the standard deviation or the effect size is different than we estimated in the adjusted sample size calculation.

Table 2. Sensitivity Analysis on Sample Size

Total Sample Size with Power = 0.80			
	$\delta = 5$	$\delta = 4.785$*	$\delta = 4.57$
$\sigma = 14.5$	272	296	324
$\sigma = 15$*	290	**318**	348
$\sigma = 15.5$	310	338	372

**The effect size and standard deviation used for adjusted sample size calculation in part iii.*

v) Interim Analysis Plan

The DSMB will meet during the trial as multiple information fractions, which is the fraction of total information expected at the scheduled end of the trial, are collected. Firstly, they will need to make safety decisions based on data so far. If by the time of the interim look, there had been a substantial evidence of the intervention is harmful for the patients, the trial needs to be terminated early to avoid patients from further risk. Secondly, the DSMB will calculate the z-scores of the

effect sizes observed and compare it to the stopping boundaries to determine whether the difference is significant enough such that the p-value is less than the total alpha accumulated by the look.

The family-wise error rate is prespecified as 0.05, and by specifying O'Brien-Fleming alpha-spending function, one can determine the rate at which the overall type I error is to be spent during the trial. The alpha will accumulate to 0.05 at the final look.

The proposed trial protocol scheduled four interim looks and one final look. The stopping rule for each look is summarized in the table below (refer to Appendix Sect.2 for PASS output).

Table 3. Interim Stopping Rules

Look	**Info**	**Lower Bndry**	**Upper Bndry**	**Inc Alpha**	**Total Alpha**	**Inc Power**	**Total Power**
1	0.2	-4.877	4.877	0.000	0	0.000	0.000
2	0.4	-3.357	3.357	0.001	0.001	0.059	0.059
3	0.6	-2.680	2.680	0.007	0.008	0.259	0.319
4	0.8	-2.290	2.290	0.017	0.024	0.289	0.608
5	1.0	-2.031	2.031	0.026	0.05	0.194	0.802

Figure 1. Plot of Stopping Boundaries

Fig.1 represents the lower boundary of the stopping rule in green and the upper boundary in red. Since the proposed trial using active-control design, two boundaries are symmetric around the horizontal axis. The x-axis shows the information fraction, and the y-axis shows the z-scores of the effect size. If the observed z-score crosses the upper boundary and its corresponding p-value is less than the cumulative alpha at the look, it indicates that before the scheduled end of trial, there is a substantial evidence that the intervention is superior to the control so that the trial

could end early; on the other hand, if the observed z-score crosses the lower boundary and its corresponding p-value is less than the cumulative alpha at the look, it indicates that before the scheduled end of trial, there is a substantial evidence that the control is superior to the intervention so that there is no need to continue the trial.

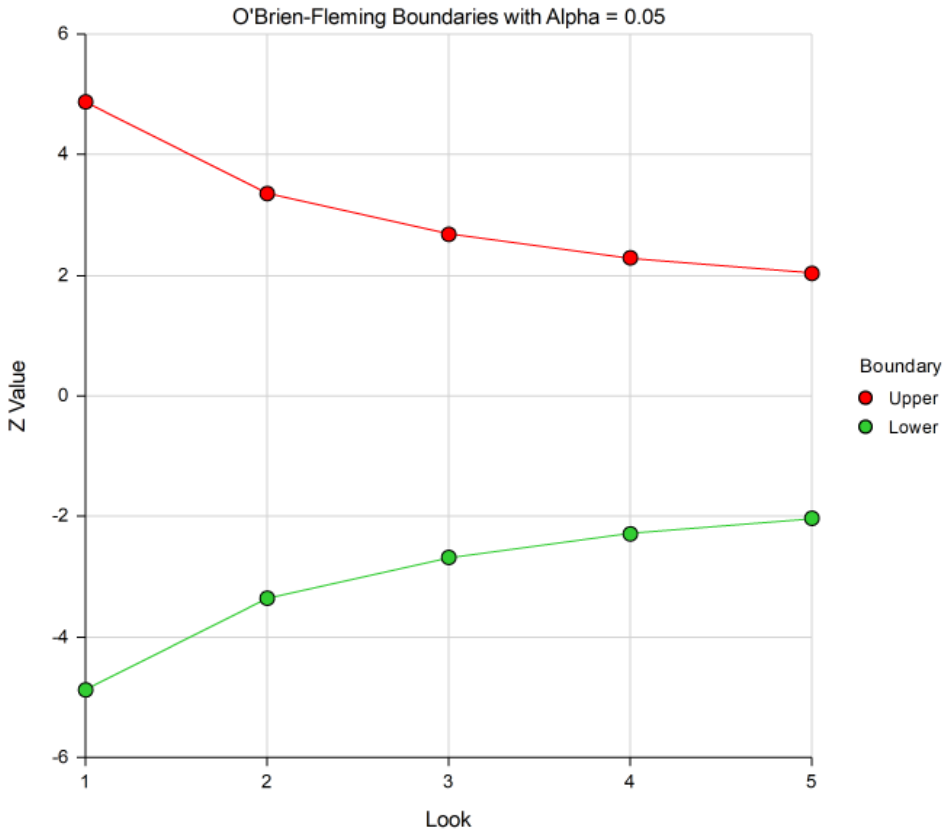


Fig.1 Stopping Boundaries for Interim Looks

6 Safety Considerations

- *Sepsis arthritis*

Sepsis arthritis a common side effect after intraarticular injections for the local treatment of knee OA. The incidence of septic arthritis of the knee after corticosteroid injection, another widely used treatment for knee OA, ranges from one in 3000 to one in 50,000[22]. Thus, it is necessary to

include it as a safety outcome since the proposed trial consists of the Sprifermin intraarticular injection as the intervention. During the trial, it is crucial to monitor any events that are potentially sepsis arthritis, as bacterial arthritis of any cause is associated with up to 15% mortality and residual impairment of joint function in up to 50% of survivors[23].

If patients report knees are showing symptoms of swollen, warm, and redness, investigators will need to consider the probability of sepsis arthritis. An accurate diagnosis will be performed by extracting the fluid sample using arthrocentesis and conducting lab evaluation. The lab will perform a white cell count on the fluid, which will be usually very high[24].

- *Musculoskeletal and connective tissue orders: decreased range of motion, knee pain, knee swelling*

Decreased range of motion, knee pain, and knee swelling are common adverse events after intraarticular injections. In addition to reporting them as adverse events, patients and investigators need to pay closer attention if the symptoms persist. These symptoms often concur with common serious adverse events following knee OA injection treatment such as arthralgia, arthritis, and joint effusion. Thereby it is important to record them and follow up with the patient if the symptoms continue after taking anti-inflammatory medications.

The range of motion of knee is measured by a goniometer. Patients will need to lie down on their back on a flat surface. To measure the knee extension, gently push the knee into the floor; to measure knee flexion, bend the knee as far as possible. Knee swelling is assessed by measuring the knee grith at level of patella. The level of knee pain can be subjective to the patient, but any abnormal knee pain that is more severe than regularly expected should be reported and recorded

as adverse events, and a more detailed examination should be conducted in case of serious adverse events.

- *Other safety outcomes*

The trial will continuously monitor other safety outcomes, such as all-cause mortality, hospitalization, prolongation of hospitalization, disability, and any other events that have a negative impact on patient's physical function or daily life activities.

7 Limitations and Late-Breaking Problems

The efficacy of the intervention arm, which is the regimen of Sprifermin and physical therapy, has not yet been studied in pre-Phase III trials. Thereby, in the proposed trial we used an estimated improvement on the primary outcome with the addition of Sprifermin for effect size and sample size calculation. The result needs to be confirmed in independent trials.

Because the individualized therapy plans are made by physical therapists from multiple clinical sites, the strategies behind the plans may be based on their various experience. Thus, there may be a differential effect on the outcome. The sample size may need to be inflate due to the potential therapists' cluster effect.

8 References

1. Cross M, Smith E, Hoy D, et al. The global burden of hip and knee osteoarthritis: estimates from the global burden of disease 2010 study. Ann Rheum Dis. 2014; 73: 1323-1330.
2. Wang Y, Nguyen U, Lane NE, et al. Knee osteoarthritis, potential mediators, and risk of all-cause mortality: data from the Osteoarthritis Initiative. Arthritis Care Res. 2021; 73: 566-573.
3. Ringdahl Erika. Treatment of knee osteoarthritis. Am Fam Physician. 2011 Jul; 83(11): 1287-1292.
4. Mora JC, Przkora R, Cruz-Almeida Y. Knee osteoarthritis: pathophysiology and current treatment modalities. J Pain Res. 2018; 11: 2189-2196.
5. Deyle GD, Allen CS, Allison SC, et al. Physical therapy versus glucocorticoid injection for osteoarthritis of the knee. N Engl J Med. 2020; 382: 1420-1429.
6. Deyle GD, Gill NW, Rhon DI, Allen CS, et al. A multicentre randomized, 1-year comparative effectiveness, parallel-group trial protocol of a physical therapy approach compared to corticosteroid injections. BMJ Open. 2016;6
7. Lohmander LS, Hellot S, Dreher D, et al. Intraarticular sprifermin (recombinant human fibroblast growth factor 18) in knee osteoarthritis: a randomized, double-blind, placebo-controlled trial. Arthritis Rheumatol. 2014 Jul; 66 (7): 1820-1831.
8. Eckstein F, Kraines JL, Aydemir A, et al. Intra-articular sprifermin reduces cartilage loss in addition to increasing cartilage gain independent of location in the femorotibial joint: post-hoc analysis of a randomised, placebo-controlled phase II clinical trial. Ann Rheum Dis. 2020 Apr; 79(4): 525-528.

9. Anwer S, Alghadir A, Zafar H, Brismee JM. Effects of orthopaedic manual therapy in knee osteoarthritis: a systematic review and meta-analysis. J. Physio. 2018 Sept. 104 (3): 264-276.
10. Eckstein F, Wirth W, Guermazi A, Maschek S, Aydemir A. Brief report: intraarticular sprifermin not only increases cartilage thickness, but also reduces cartilage loss: location-independent post hoc analysis using magnetic resonance imaging. Arthritis Rheumatol. 2015 Nov;67(11):2916-22.
11. Abbott JH, Chappe CM, Fitsgerald GK, et al. The incremental effects of manual therapy or booster sessions in addition to exercise therapy for knee osteoarthritis: a randomized clinical trial. JOSPT. 2015 Nov; 45(12): 975-983.
12. Kolahi, J., Bang, H. and Park, J. (2009), Towards a proposal for assessment of blinding success in clinical trials: up-to-date review. Community Dentistry and Oral Epidemiology, 37: 477-484.
13. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt L. Validation study of WOMAC: a health status instrument for measuring clinically important patient-relevant outcomes following total hip or knee arthroplasty in osteoarthritis. J Orthop Rheumatol 1988; 1: 95–108.
14. McConnell S, Kolopack P, Davis AM. The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC): a review of its utility and measurement properties. Arthritis Rheum. 2001 Oct;45(5):453-61.
15. Brazier JE, Harper R, Jones NM, et al. Validating the SF-36 health survey questionnaire: new outcome measure for primary outcome. British Medical Journal. 1992; 305:160.

16. W. Wirth and F. Eckstein, "A Technique for Regional Analysis of Femorotibial Cartilage Thickness Based on Quantitative Magnetic Resonance Imaging," IEEE Transactions on Medical Imaging. 2008 Jun; 27(6): 737-744.
17. Lohmander LS, Hellot S, Dreher D, Krantz EF, Kruger DS, Guermazi A, et al. Intraarticular sprifermin (recombinant human fibroblast growth factor 18) in knee osteoarthritis: randomized, double-blind, placebo-controlled trial. Arthritis Rheumatol 2014; 66:1820–31.
18. Dobson F, Hinman RS, Hall M, et al. Reliability and measurement error of the osteoarthritis research society international (OARSI) recommended performance-based tests of physical function in people with hip and knee osteoarthritis. Osteoarthritis and Cartilage. 2017 Nov; 25(11): 1792-1796.
19. Ateef M, Kulandaivelan S, Tahseen S. Test-retest reliability and correlates of 6-minutes walk test in patients with primary osteoarthritis of knees. Indian Journal of Rheum. 2016; 11(4): 192-196.
20. Fitzgerald GK, Fritz JM, Childs JD, et al. Exercise, manual therapy, and use of booster sessions in physical therapy for knee osteoarthritis: a multi-center, factorial randomized clinical trial. Osteoarthritis and Cartilage. 2016 Aug;24(8): 1340-1349.
21.
22. Munigangaiah S, O'Sullivann TA, and Lenehan B. Simultaneous bilateral septic arthritis of the knee after intraarticular steroid injection: a clinical report. J Nat Sci Biol Med. 2014 Jul; 5(2):485-487.

23. Shemesh S, Heller S, Salai M, et al. Septic arthritis of the knee following intraarticular injections in elderly patients: report of six patients. Isr Med Assoc J. 2011 Dec;13(12):757-760.
24. Chan BY, Crawford AM, Kobes PH, et al. Septic arthritis: an evidence-based review of diagnosis and image-guided aspiration. AJR.2020 Sept; 215(2):568-581.

Appendix

Section 1. Adjusted Effect Size Calculation

	Intervention Group (Sprifermin + physical therapy)	Control Group (Physical therapy)
% Crossover	5%	5%
% Noncomplier	15%	15%
% Full complier	85%	85%

The proposed trial obtains the mean WOMAC change from baseline and corresponding standard deviation (SD = 11.8) from previously published study on the efficacy of physical therapy on knee OA and set it as the mean WOMAC score change for the control group. The change is in the negative direction and take the absolute value of it. Thus,

$$\mu_C = 33.9$$

No previous study has been conducted on the same arms as the proposed trial. Thereby, we estimate that Sprifermin will brings a 20% improvement on physical therapies. Thus,

$$\mu_I = 33.9 \times 1.2 = 40.68$$

The unadjusted effect size is,

$$\mu_I - \mu_C = 6.78$$

We estimate the mean WOMAC change for noncompliers is the average of the intervention and control groups,

$$\mu_{Non-complier} = \frac{(\mu_I + \mu_C)}{2} = 37.29$$

For the intervention group, the adjusted mean outcome is,

$$\mu_{adj\_I} = 0.05 \times 33.9 + 0.15 \times 37.29 + 0.80 \times 40.68 = 39.8325$$

For the control group, the adjusted mean outcome is,

$$\mu_{adj\_C} = 0.05 \times 40.68 + 0.15 \times 39.29 + 0.80 \times 33.9 = 35.0475$$

The adjusted effect size, which takes into account of noncompliance and crossover is,

$$\mu_{adj\_I} - \mu_{adjC} = 4.785$$

PASS calculation for *adjusted* sample size:

Group-Sequential Tests for Two Means (Legacy)

Numeric Results for Two-Sided Hypothesis Test of Means

Target Power	Actual Power	N1	N2	N	Mean1	Mean2	S1	S2	Alpha
0.8	0.80228	159	159	318	39.8325	35.0475	15	15	0.05

The *adjusted* sample size is 318.

Section 2.

Interim Stopping Rule –

Details when Spending = O'Brien-Fleming, N1 = 159, N2 =159, S1 = 15, S2 = 15, Diff = 4.785

Look	Time	Info	Lower Bndry	Upper Bndry	Nominal Alpha	Inc Alpha	Total Alpha	Inc Power	Total Power
1	0.2	0.2	-4.87688	4.87688	0.00000	0.00000	0.00000	0.00016	0.00016
2	0.4	0.4	-3.35695	3.35695	0.00079	0.00079	0.00079	0.05946	0.05962
3	0.6	0.6	-2.68026	2.68026	0.00736	0.00683	0.00762	0.25915	0.31877
4	0.8	0.8	-2.28979	2.28979	0.02203	0.01681	0.02442	0.28945	0.60822
5	1.0	1.0	-2.03100	2.03100	0.04226	0.02558	0.05000	0.19406	0.80228

Drift 2.84429