Which Exercise for Low Back Pain? (WELBack) trial predicting response to exercise treatments for patients with low back pain: a validation randomised controlled trial protocol

Luciana G Macedo, Paul W. Hodges, Geoff Bostick, Mark Hancock, Maude Laberge, Steven Hanna, Greg Spadoni, Anita Gross, Julia Schneider

ABSTRACT

Introduction Exercise therapy is the most recommended treatment for chronic low back pain (LBP). Effect sizes for exercises are usually small to moderate and could be due to the heterogeneity of people presenting with LBP. Thus, if patients could be better matched to exercise based on individual factors, then the effects of treatment could be greater. A recently published study provided evidence of better outcomes when patients are matched to the appropriate exercise type. The study demonstrated that a 15-item questionnaire, the Lumbar Spine Instability Questionnaire (LSIQ), could identify patients who responded best to one of the two exercise approaches for LBP (motor control and graded activity). The primary aim of the current study will be to evaluate whether preidentified baseline characteristics, including the LSIQ, can modify the response to two of the most common exercise therapies for non-specific LBP. Secondary aims include an economic evaluations with a cost-effectiveness analysis.

Methods and analysis Participants (n=414) will be recruited by primary care professionals and randomised (1:1) to receive motor control exercises or graded activity. Participants will undergo 12 sessions of exercise therapy over an 8-week period. The primary outcome will be physical function at 2 months using the Oswestry Disability Index. Secondary outcomes will be pain intensity, function and quality of life measured at 2, 6 and 12 months. Potential effect modifiers will be the LSIQ, self-efficacy, coping strategies, kinesiophobia and measures of nociceptive pain and central sensitisation. We will construct linear mixed models with terms for participants (fixed), treatment group, predictor (potential effect modifier), treatment group×predictor (potential effect modifier), physiotherapists, treatment group×physiotherapists and baseline score for the dependent variable.

Ethics and dissemination This study received ethics approval from the Hamilton Integrate Research Ethics Board. Results will be submitted for publication in peer-reviewed journals.

Trial registration number NCT04283409.

Strengths and limitations of this study

- This study will be a randomised controlled trial (RCT) fully powered to address an effect modification analysis.
- This replication study is a critical step to test the robustness of findings from a previous preplanned effect modification analysis in a high-quality RCT.
- A cost-effectiveness analysis will allow for better clinical decision making on the selection of the most appropriate form of exercise.
- Blinding of patients and therapists is not possible within an exercise trial and is a limitation of this study.
- Out-of-pocket costs will be self-reported and may suffer from recall bias.

INTRODUCTION

Exercise therapy is the most commonly endorsed treatment for chronic non-specific low back pain (LBP) in clinical guidelines and systematic reviews. Although exercise is more effective than no intervention, the effect sizes of exercise, like other back pain treatments, are small to moderate, and no exercise approach (eg, Pilates or general exercises) has been shown to be superior to another exercise. The most commonly cited reason for the relatively small average effect size for exercise is the heterogeneity of people presenting with LBP. Improved precision in matching patients to specific types of exercise (ie, a personalised medicine approach to exercise) could increase treatment effects.

Recent progress has been made in the triage of individuals attending primary care for LBP. The STarT Back Tool uses risk stratification to aid selection of those who would benefit from physiotherapy, primarily
exercise for a medium-risk group, or exercise plus more complex psychosocial interventions for a high-risk group. Although this tool is more effective and cost-effective than usual care,\(^9\)\(^10\) this stratification care model does not provide guidance regarding the type of exercise to provide to individuals with LBP. Those who have undergone risk stratification according to the STarT Back tool and require exercise therapy (~75%)\(^11\) are still faced with the decision of ‘Which exercise will produce the best outcomes?’ Identification of subgroups of responders to treatment, including different exercise types, has been one of the leading research priorities in the last two international consensus statements on priorities for LBP research.\(^12\)\(^13\)

There has been some effort towards derivation of clinical decision support tools for LBP.\(^14\)\(^17\) Most tools were not developed using strict methodological criteria, such as the evaluation of plausible effect modifiers within a randomised controlled trial (RCT) or were derived from studies that were underpowered to evaluate interaction effects.\(^18\) Further, these tools were not subsequently validated or tested for effectiveness within an RCT design. Validation of clinical decision support tools is critical as derivation studies are highly prone to spurious findings. Multiple systematic reviews of clinical support tools conclude that we do not need more development studies but rather validation studies.\(^15\)\(^17\)\(^19\) This is an important challenge within the LBP literature where many promising studies are neither replicated nor translated into clinical practice. Validation studies are essential before recommending a change in policy and clinical practice.\(^19\)

A recently published preplanned secondary analysis of an RCT investigated whether clinical characteristics could identify patients who benefit more from either motor control exercises or a graded activity programme using principles of cognitive behavioural therapy.\(^20\)\(^22\) Both exercise approaches have evidence of effectiveness from high-quality systematic reviews\(^23\)\(^26\) but no evidence for superiority when tested in heterogeneous populations.\(^21\)\(^25\) The study results suggest that a simple 15-item self-report questionnaire, the Lumbar Spinal Instability Questionnaire (LSIQ), that asks questions primarily related to tasks that provoke and relieve symptoms,\(^27\) could identify patients who responded best to either motor control exercise or graded activity. As hypothesised, individuals with LBP and a low LSIQ score (<9 points) had a statistically and clinically significant better outcome when receiving graded activity than motor control exercise 1 year postintervention. Conversely, individuals with a high LSIQ score (≥9 points) achieved better outcomes with motor control exercise than those prescribed graded activity. These results provided preliminary evidence that simple and easy to collect measures might enable better matching of exercise and significantly improve outcomes for patients with LBP. It is only after validation that these results may be implementable in clinical practice.

Guidelines for the investigation of subgroups of respondents to treatment recommend the use of a small number of plausible potential effect modifiers.\(^18\) In the aforementioned derivation study,\(^22\) theoretical models behind potential mechanisms of action of the interventions were used to derive potential effect modifiers.\(^22\) Motor control exercises use a motor learning approach to address features of trunk posture, movement and muscle activation that are related to symptoms. This intervention should work best on those who lack neuromuscular coordination or have mechanically related problems. Graded activity uses principles of cognitive behavioural therapy to address LBP-related beliefs and fears. This intervention should work best for individuals who are deconditioned and have fear avoidance and other unhelpful beliefs and attitudes towards their pain.

This study will also include an economic evaluation, including cost-effectiveness and cost-utility analyses, and take into consideration the LSIQ scores. The interventions delivered as part of this study have the same direct delivery cost in that they require the same resources. However, there is some evidence that patients receiving graded activity intervention may seek less additional care for their LBP than those receiving motor control.\(^26\) Theoretically, traditional graded activity interventions include a large self-management component that reduces the need for seeking additional care.

The primary aim of this study is to evaluate whether the score on the LSIQ modifies the response of patients with chronic non-specific LBP to motor control exercises or graded activity. The secondary aim is to compare the cost-effectiveness between patients who did and did not receive the ‘matched’ intervention at 12 months of follow-up. An exploratory aim is to evaluate whether other effect modifiers contribute to the identification of subgroups of patients that may best respond to motor control exercises or graded activity.

**METHODS AND ANALYSIS**

**Study design**

The proposed study is an RCT with effect modification and economic evaluations (figure 1). This protocol was structured based on the Standard Protocol Items: Recommendations for Interventional Trials: SPIRIT Checklist for protocols of RCTs.\(^28\)

**Participant recruitment**

Participants will be recruited by primary care professionals (eg, family physicians and physiotherapists) working in Ontario, Canada. The primary care professional will assess the participant for eligibility and send referral forms to the research team. The research team will confirm eligibility and acquire informed consent prior to the participant’s inclusion in the study. See consent form in the online supplemental material.

**Participants**

**Inclusion criteria**

Patients will be eligible for inclusion if they meet all of the following inclusion criteria:
► Chronic non-specific LBP (>3 months) with or without leg pain.
► Back pain being the primary musculoskeletal complaint of the patient.
► Aged between 18 and 80 years.
► Adequate English to allow response to questionnaires and communication with a physiotherapist.
► Moderate or greater pain or disability measured using questions 7 and 8 of the 36-Item Short Form Survey.30
► Moderate- or high-risk classification on the STarT Back Tool, which indicates appropriateness of physiotherapy and thus an exercise programme.9

Exclusion criteria
Participants will be excluded if they meet at least one of the following exclusion criteria:
► Nerve root compromise (two of strength, reflex or sensation affected for the same nerve root).
► Suspected or confirmed serious pathology (eg, infection, fracture, cancer, inflammatory arthritis or cauda equina syndrome).
► Current pregnancy.
► Scheduled or on the wait list for surgery during treatment and follow-up periods.
► Cognitive impairment that precludes participant from consenting, completing study questionnaires or complying with exercise recommendations (eg, dementia or Alzheimer’s disease).
► Severe neuromuscular condition (eg, spinal cord injury) that precludes participant from engaging in active exercise.
► Clinical assessment, by a family physician, indicating that the participant is not suitable for active exercises.

The Physical Activity Readiness Questionnaire may be used as a guide to refer for clinical assessment.31

Treatment randomisation
Participants will be randomised into treatment groups using a REDCap randomisation module, a user-friendly, comprehensive and secure randomisation service for clinical trials that is run on the internet. This is a remote administered method to allocate interventions to participants. Randomisation will be blocked and stratified by physiotherapist and by the score on the LSIQ questionnaire (high, >9, and low, ≤9). A statistician outside of the study will generate the randomisation sequence and upload it to REDCap. Following consent, and at the time of the first treatment sessions, the treating physiotherapists will log into the website and receive information regarding the randomised treatment allocation. This process will guarantee allocation concealment.

Blinding
The study will be randomised with concealed allocation and assessor blinding. Blinding of therapists and participants is generally not possible in studies of physical interventions (ie, exercise).32 However, therapists and participants will be blinded to the baseline results of the primary moderator, the LSIQ. In addition, at the end of the first treatment session, participants will complete the Treatment Credibility Questionnaire to control for their expectations towards treatment.

Study interventions
This study is a comparison of two common interventions, motor control exercises and graded activity, delivered by physiotherapists for the management of LBP. These two interventions have received substantial attention over the past decades as they have been developed based on evidence-based theoretical models. Both groups will include the delivery of education on pain neurophysiology, education, ergonomics and advice to stay active.1 26 33

Participants in each group will receive 12 individually supervised sessions of approximately half an hour (as per usual clinical practice) over an 8-week period (two sessions per week in the first 4 weeks and one session per week in the second 4 weeks).34 The treatment sessions are designed to become less frequent and to promote transition to independent exercise and may become shorter in time as they progress. Participants included in both exercise programmes will be advised to complete home exercise daily during the programme and after discharge. Participants will be asked to refrain from receiving other interventions such as chiropractic treatment, massage therapy and medication, although this will not preclude from patient participation. We will collect information on cointerventions on both groups.

Treatment will be provided by 22 physiotherapists at 15 different locations in Ontario, Canada. All physiotherapists have been trained to deliver both interventions.

Macedo LG, et al. BMJ Open 2021;11:e042792. doi:10.1136/bmjopen-2020-042792

Figure 1 Consort flow diagram. LSIQ, Lumbar Spine Instability Questionnaire.
through a 2-day workshop with Professor Paul Hodges (motor control exercise) and a 2-day workshop with Dr Geoff Bostick (graded activity), as well as the use of a treatment manual and refresher webinars. Refresher sessions (webinar) are planned each 6 months to guarantee continued compliance with the protocol. To help clearly differentiate the two interventions and decrease potential overlap, a table with key treatment components was developed (Table 1).

Motor control exercise

Motor control exercise was developed within a mechanical model of LBP. It is based on a large body of evidence suggesting that individuals with LBP have impaired control and coordination of the trunk muscles, including the deep trunk muscles (eg, transversus abdominis and multifidus). The underlying premise is that for many with chronic LBP, ongoing nociceptive input from the periphery secondary to suboptimal tissue loading contributes to the ongoing pain. The primary goal of motor control exercises is to retrain optimal control and coordination of the spine and limbs.

The motor control exercise will be based on the programme reported by Hodges et al and similar to the protocol previously used in the initial study that this trial replicates and extends and a Canadian feasibility study (unpublished). The intervention is called motor control as it uses principles of motor learning, such as segmentation, simplification and task-specific practice to retrain control of trunk muscles activation, body alignment and movement. The intervention is based on assessment of the individual patient’s motor control features, their relationship to symptoms and the patient’s individual treatment goals (set collaboratively with the physiotherapist).

The first stage of the treatment involves assessment of the postures, movement patterns and muscle activation associated with symptoms and implementation of a retraining programme. The programme is designed to improve activity of muscles assessed to have poor control (commonly the deeper muscles such as transversus abdominis, multifidus, pelvic floor and diaphragm) and to reduce the activity of any muscle identified to be overactive, commonly the large, more superficial trunk muscles such as obliquus externus abdominis that the individual may use to brace and guard the trunk.

Participants are taught how to contract/control these muscles independently from the superficial trunk muscles and progress until the patient is able to maintain independent contractions of the target muscles while maintaining normal respiration. Exercises for breathing control, control of posture and movement of spine and limbs are individualised based on the presentation of the patient.

The second stage of the treatment involves progression of the exercises towards functional activities using a combination of static and dynamic tasks. Throughout this process, the recruitment of the trunk muscles, posture, movement pattern and breathing are assessed and modified as necessary.

| Principles                        | Graded activity | Motor control |
|-----------------------------------|-----------------|---------------|
| Goal setting                      | ✓               | ✓             |
| Pain contingent                   | X               | ✓             |
| Time contingent                   | ✓               | X             |
| Quotas/pacing                     | ✓               | X             |
| Reinforce well behaviour and address illness behaviour | ✓ | X |
| Education regarding pain system   | ✓               | ✓             |
| Reassurance                       | ✓               | ✓             |
| Generalised (whole body) exercises with consideration of specific muscle activity | X | ✓ |
| Generalised (whole body) exercises without consideration of specific muscle activity | ✓ | X |
| Specific (localised) exercises    | X               | ✓             |
| Correction of activation of muscles | X           | ✓             |
| Correction of posture             | X               | ✓             |
| Strength training                 | ✓               | ✓             |
| Cardiovascular/fitness training   | ✓               | ✓             |
| Coordination training             | X               | ✓             |
| Correction of motor patterns      | X               | ✓             |
| Muscle stretching                 | ✓               | ✓             |
| Breathing pattern                 | X               | ✓             |
| Consideration of continence       | X               | ✓             |
| Correction of provocative movement patterns | X | ✓ |
| Relaxation techniques             | X               | ✓             |
| Progress to functional activities | ✓               | ✓             |
| Use feedback (eg, ultrasound, electromiography and biofeedback) to enhance learning of movement pattern or muscle activation | X | ✓ |
| Home exercises                    | ✓               | ✓             |
| Psychologically informed          | ✓               | X             |
| Mechanically informed             | X               | ✓             |

Table 1: Key intervention principles

Graded activity

Graded activity is based on the biopsychosocial model. In this model, disability is determined not only by the underlying pathology, but also mostly by contextual factors such as social, cognitive, emotional and behavioural factors.

The primary goal of graded activity is to address individual modifiable contextual factors associated with the activity-related pain experience such as self-efficacy, pain-related fear, fear of movement and unhelpful beliefs/behaviours about back pain while at the same time addressing physical impairments such as endurance, muscle strength and balance.
The graded activity exercise programme is based on the treatment programme reported by Lindström et al. and similar to the protocol previously used in the initial study that this trial replicates and extends, and a Canadian feasibility study (unpublished). A primary goal of the programme is to increase activity tolerance by performing individualised and submaximal exercises in addition to addressing illness behaviours and reinforcing well behaviours. The programme is based on meaningful activities that each patient identifies as problematic as a result of their LBP. The activities in the programme are progressed in a time-contingent manner (rather than a traditional pain-contingent manner) from the baseline assessment ability to a target goal set jointly by patient and physiotherapist.

Participants receive daily quotas and are instructed to only perform the agreed amount, not less or more, even when they feel they are capable of doing more. Cognitive–behavioural therapy principles, such as reinforcement and reassurance, are used to help participants overcome the natural reluctance to engage in activities and exercise while experiencing pain.

**Data collection**

Outcomes (described further) will be assessed at similar times for both exercise groups at the follow-ups of 2, 6 and 12 months. All core outcomes in LBP are self-reported and therefore will be collected using a REDCap URL link sent to participants via email. If the participant does not complete the questionnaire within 2 days, they will be sent a reminder, and if they still do not respond, they will be contacted via phone. Alternatively, if the participant does not own an email account or does not have access to the internet, the questionnaires will be completed over the phone by an investigator unaware of treatment allocation.

Pain pressure threshold is the only outcome measure that will be collected in person as part of this study. Baseline assessment will be performed by the treating physiotherapist after consent and prior to randomisation. At the end of the treatment prior to discharge, pain pressure threshold will be assessed again, but at this time, physiotherapists will no longer be blinded to treatment allocation.

**Baseline assessment**

Demographics characteristics will be collected at baseline using self-report questionnaires. In addition, a comorbidity questionnaire will be collected at baseline to identify potential conditions that might confound the results of this study, such as the presence of fibromyalgia or other musculoskeletal conditions.

**Outcome measures**

**Primary outcome**

Disability at 2 months

The Oswestry Disability Index (ODI) is a disease-specific disability questionnaire with 10 items (total score from 0 to 100). This questionnaire is one of the most used in LBP populations and is a core outcome measure recommended to evaluate pain-related disability in this patient group. The ODI has evidence for validity, reliability and responsiveness in the LBP population.

**Secondary outcomes**

**Disability**

ODI collected at 6 and 12 months of follow-up.

**Function**

Function will be assessed using the Pain-Specific Functional Scale (PSFS). The PSFS requires participants to identify three problematic activities and to rate these activities on a scale from 0 to 10. The total score is from 0 to 10 as an average of the three activities. This scale has evidence for validity, reliability and responsiveness in an LBP population.

**Pain**

Average pain intensity over the last week will be measured using a Numerical Pain Rating Scale from 0 to 10 anchored with ‘no pain’ at 0, and ‘worst pain imaginable’ at 10. This measure is one of the most widely used to assess pain in LBP studies and is also a core outcome measure in the clinical assessment of LBP.

**Quality of life**

Health-related quality of life will be assessed using the Five-Level EuroQol Five-Dimensional (EQ-5D-5L). The EQ-5D-5L-derived health utilities will be used for a cost-utility analysis. The EQ-5D-5L can define a total of 3125 health states. It can be converted to a utility index anchored at 0 for a state being equal to death and 1 for full health using the Canadian scoring algorithm. The utility index, combined with life expectancy, allows for the estimation of quality-adjusted life years (QALY). The QALY is a recommended generic outcome measure in health economic evaluation in Canada. The EuroQol Five-Dimensional is the most commonly used health-related quality of life measure in LBP cost-effectiveness analysis.

‘IMPACT of LBP’ at 12 months: Nine-Item Patient-Reported Outcome Measurement Information System (PROMIS)

This outcome is recommended by the recent National Institutes of Health (NIH) task force report on research standards for LBP. The outcome combines nine items from the Patient-Reported Outcome Measurement Information System short form widely recommended in consensus documents on outcome measures for LBP. IMPACT of LBP as defined by NIH covers the domains of pain intensity, pain interference with normal activities and functional status. A score from 8 (least impact) to 50 (greatest impact) is produced. Although this measure is relatively new, all included items have undergone substantial psychometric testing and are used in the field. The key advantages of this outcome are that it focuses on the impact of LBP and is multidimensional. The questionnaire has greater sensitivity to change than the commonly

Macedo LG, et al. BMJ Open 2021;11:e042792. doi:10.1136/bmjopen-2020-042792
used Roland Morris Disability Questionnaire and change scores correlate more strongly with patient satisfaction.75

**Monitoring variables**

**Adverse events**

The occurrence of adverse events (eg, muscle soreness and transient increase in LBP levels) will be monitored by the physiotherapists during the intervention and using open-ended questions during the predefined data collection follow-ups.

**Adherence**

Adherence with the exercise programme will be assessed using patient treatment files that will be collected from physiotherapists at the end of the study. In addition, participants will be asked to keep a log of the home exercise programme for the duration of the intervention. Adherence with home exercises after discharge will be measured using an ordinal scale (How often did you perform home exercises? None of the time, some of the time, most of the time, almost all of the time or all of the time) at each study follow-up.

**Cointerventions**

Participants will be asked to report any type of cointerventions (eg, chiropractic or massage therapy) during each predefined data collection follow-ups, which will allow for the observation of cross-overs. Pain medication will be permissible as it would be unethical to withhold medication. However, this information will be collected for both groups.

**Potential effect modifiers**

Effect modifiers will be collected at baseline and at 2 and 12 months of follow-up. Baseline data will be used for the effect modification analysis. Twelve-month data will be used for a secondary analysis of mediators of outcomes.

**Lumbar Spine Instability Questionnaire**

The primary effect modifier is used to evaluate the construct of self-reported ‘clinical instability’. The questionnaire was developed from domains identified by an expert consensus project of clinical features of ‘instability’.27 It has face validity, good internal consistency and22 moderate to high test–retest reliability (intra-class correlation coefficient (ICC)=0.75).77 78 A recently published Rasch analysis demonstrated the questionnaire to be unidimensional.79 The measure has not been validated against any objective measure of ‘instability’, and there is discussion whether this questionnaire reflects other mechanisms, such as underlying physiological pain mechanism.43

**Orebro LBP Screening Questionnaire**

Used to identify clinical characteristics, symptom severity, attitudes and behaviours towards LBP.85 The questionnaire has been found to be a prognostic indicator in LBP and to have good test–retest reliability.11 80–84

**Tampa Scale of Kinesiophobia**

Used to evaluate fear of movement and attitudes and behaviour towards pain.85

**Coping strategies questionnaire**

This is used to evaluate beliefs and behaviours towards LBP.86 The questionnaire has been shown to have good concurrent validity and test–retest reliability.87–89

**painDetect**

This is a screening questionnaire to identify features of neuropathic pain that has been used to distinguish between nociceptive and neuropathic pain components in LBP.86 The questionnaire has been validated for use in the LBP population and has good test–retest reliability and internal consistency.91 92

**Smart clinical criteria checklist**

This is used to discriminate between individuals with a primary mechanism for persistence of pain explained by nociceptive, peripheral neuropathic or central sensitisation (nocicplastic) mechanisms.93

**Pain pressure threshold (PPT)**

PPT will be measured with an electronic pressure algometer (stimulation surface area of 1 cm²) at the lumbar spine (the point with the most pain at the lumbar spine—indicative of local hyperalgesia) and at distant site (thumbnail on the opposite side of the body—indicative of generalised hyperalgesia).94 The lumbar spine measure will be performed with the participant in side lying and hips and knee at approximately 45° of flexion. Thumbnail measures will be performed with the participants sitting with the thumb resting over a hard surface. PPT will be recorded as the point which the sensation of pressure changes to a sensation of pain. Pressure will be increased from 0 N at 40 N/s until the participants indicate that they have reached their PPT. Two consecutive measurements will be performed at each site, with 10 s recovery between repeated applications. The threshold pressure will be transformed to kilogram per square centimetre.

**Depression: CES-D scale**

This is used to evaluate the level of anxiety and depression. This will be completed based on evidence that central pain mechanisms may interact with depressive symptoms.95

**Pain Catastrophising Scale (PCS)**

Pain catastrophising affects how people experience pain.96 The PCS is a 13-question self-report questionnaire that evaluates three domains (ruminator, magnification and helplessness). Each item is scored on a scale from 0 to 4 with the scale total score ranging from 0 to 52. A total PCS score of 30 represents clinically relevant levels of severe catastrophising in varying presentations of persistent pain.96
Central Sensitisation Inventory (CSI)
This is used to identify the presence of central sensitisation that may contribute to maintenance of pain via nociceptive mechanisms. The CSI is a two-part questionnaire designed to assess the potential for central sensitisation within the clinic. Part A includes 25 questions about symptom presentation, and part B has 10 questions about specific conditions associated with central sensitisation that have been diagnosed by a medical doctor. A cut-off score of 40 on part A of the questionnaire predicts presence of central sensitisation in individuals with persistent pain.

Gender
Some evidence suggests that rehabilitation programmes such as behavioural therapy might require different implementation strategies based on gender. Given that graded activity includes a behavioural component, we will include gender as a potential effect modifier to identify if response to exercise differs between gender groups.

Physiotherapist perception
To investigate whether physiotherapists’ perceptions align with theoretical frameworks, therapists will complete a Likert scale from 0 to 10, where 0 is graded activity and 10 is motor control. They will be asked to place each patient on the scale according to their perception regarding which intervention would be most appropriate for each individual patient. This assessment will be completed at the end of the first assessment/treatment session.

Cost-effectiveness measures
We will analyse costs from the healthcare system perspective, societal perspective and patient perspective.

Healthcare system perspective
Costs from the Ontario Health Insurance Plan will be collected using health administrative databases at the Institute for Clinical Evaluative Sciences. We will extract and analyse data on patients’ use of all publicly funded health services for the study intervention period (12-month period) and at baseline with the 12-month period prior to the study period. Such services include visits to primary care physicians, to the emergency department, hospital admissions, rehabilitation, home care, drugs and lab tests. Patient-level cost algorithms have been developed and will enable estimating costs at the patient level.

Societal and patient perspectives
Societal costs and costs incurred by patients will be collected using a validated standardised questionnaire (Costs for Patients Questionnaire (CoPaQ)) at baseline and at 2, 6 and 12 months of follow-up (M Laberge, under review 2020) In this tool, participants provide information about costs incurred because of LBP, including medical costs directly related to service use (such as what they paid out-of-pocket for their medications) and non-medical costs (such as transportation to an appointment), in a defined period. In addition, participants will enter indirect costs (such as income loss) for themselves and for their family members as it relates to their LBP. These will be used to estimate costs from the societal perspective and from the patient perspective. The CoPaQ will enable the inclusion of indirect costs into the economic evaluations, as well as the evaluation of costs from the patient perspective. Thus, societal perspective analysis will include healthcare system costs, as well as indirect costs from loss of productivity.

Sample size calculation
It has been demonstrated through simulation studies that the sample size of a 2-x-2 interaction (eg, treatment x-effect modifier) in a mixed effects model is fourfold of the sample to detect a main effect of the same magnitude. It is argued that interaction effects are usually smaller in validation studies than pilot studies, and therefore, the study was powered for a clinically significant interaction of 15 points. Assuming from previous systematic reviews the main effect of these exercise approaches compared with no treatment is approximately 15 points on a 100-point scale for the ODI at 2 months, an interaction of 15 would mean that the effect would be approximately 22.5 when participants receive the correct targeted intervention and 7.5 when participants do not receive the targeted intervention. This would generate a significant improvement in outcomes that are clearly considered clinically relevant (>10 points). Sample size was calculated for the primary outcome of function (ODI) using PASS V.16 software for mixed models tests for two means based on the methods of Vieron and Giraud. A sample size of 90 achieves 85% power to detect a difference of 15 between the two means (main effect) when the SD of the response variable is 25 and the ICC (ρ) is 0.10 using a test with a significance level of 0.05. A sample size of 360 (90×4) is required to appropriately power the study for the interaction effect. A total of 424 participants (212 per treatment group) will be included to account for a 15% loss to follow-up.

Data integrity
Data will be directly collected into the REDCap programme. Any inconsistencies in the data will be explored and resolved. The database will be maintained on a secure network that is compliant to the International Conference on Harmonisation Guideline for Good Clinical Practice. Only study personnel will have access to the password-protected database.

Retention of documents
Investigators will maintain adequate and accurate records to fully document the conduct of the study and enable verification of the study data. Study data will be archived by McMaster University for a minimum of 7 years.

Statistical analysis
Primary analysis
The primary analyses will be by intention to treat, and participants will be analysed in the groups in which they were randomised regardless of cross-over of treatment discontinuation. For the primary and secondary
outcomes, a p value of <0.05 will be considered statistically significant. Group allocation will remain masked until analyses and interpretation are finalised.

**Treatment effectiveness and effect modification analysis**

The effect modification analysis will be conducted for the primary outcome and secondary outcomes separately. Linear mixed models with terms for participants (fixed), treatment group, treatment group×time (for the effectiveness analysis), predictor (potential effect modifier such as LSQ), treatment group×predictor (for the effect modification analysis), predictor (potential effect modifier such as baseline). Linear mixed models with terms for participants (fixed), treatment group, treatment group×predictor (for the effect modification analysis), predictor (potential effect modifier such as baseline). The effect modification analysis will be conducted for the primary outcome and secondary outcomes separately.

**Economic evaluations**

We will conduct two economic evaluations along with the RCT: a cost-effectiveness analysis and a cost-utility analysis following the methodology outlined in Drummond et al. The cost-effectiveness analysis will use the primary outcome, that is, changes in the score on ODI, as the effect and the cost-utility analysis will use QALYs, derived from data collected with the EQ-5D-5L. An incremental cost-effectiveness ratio (ICER) will be calculated from the perspective of the Ministry of Health for the main outcome (ODI). We will also estimate the ICER for the cost-utility analysis with QALYs. Analyses will be conducted at the patient level. Table 2 presents a summary of analyses.

The steering committee will be composed of the primary investigator and coinvestigators, which include two knowledge users. The committee will meet two times a year to discuss the study. We will not have a data safety and monitoring committee as adverse events are not a factor of concern in this study.

**Patient and public involvement**

The study protocol was developed in collaboration with community physiotherapists with extensive experience in treating individuals with LBP.

**ETHICS AND DISSEMINATION**

**Ethics approval and trial registration**

This study received ethics approval from the Hamilton Integrate Research Ethics Board (HiREB #7986). Trial was registered online (clinicaltrials.gov).

**Dissemination of findings**

An article with the results of the effectiveness analysis, another article with the results of the cost-effectiveness analysis will be published in peer-reviewed journals. On completion of the trial, and after publication of the primary manuscripts, data requests can be submitted to the primary investigator (LM).

| Table 2 Cost-effectiveness analysis |
|-------------------------------------|
| Economic evaluation | Cost-effectiveness analysis | Cost-utility analysis |
| Outcome | Difference between ODI scores at baseline and at 12 months | Difference between in QALYs derived from EQ-5D-5L scores at baseline and at 12 months |
| Costs | ▶ MOH perspective | ▶ All healthcare costs covered by the MOH from IC/ES. |
| ▶ Socioeconomic perspective | ▶ All healthcare costs covered by the MOH+indirect costs from loss of productivity. |
| ▶ Patient perspective | ▶ Cost-sharing for healthcare services+opportunity costs. |

Incremental cost-effectiveness ratio

\[ \frac{\Delta C}{\Delta E}, \text{ where} \]

\[ \Delta C = (C_{t12} - C_{t0}) - (C_{c,t12} - C_{c,t0}) \]

\[ \Delta E = (E_{t12} - E_{t0}) - (E_{c,t12} - E_{c,t0}) \]

with:

\[ C_{l,t0} = \text{healthcare cost at 12 months of patients in the intervention group} \]

\[ C_{c,t0} = \text{healthcare cost at baseline of patients in the intervention group} \]

\[ C_{l,t12} = \text{healthcare cost at 12 months of patients in the control group} \]

\[ C_{c,t12} = \text{healthcare cost at baseline of patients in the control group} \]

\[ U_{l,t0} = \text{EQ-5D-5L score at 12 months of patients in the intervention group} \]

\[ U_{c,t0} = \text{EQ-5D-5L score at baseline of patients in the intervention group} \]

\[ U_{l,t12} = \text{EQ-5D-5L score at 12 months of patients in the control group} \]

\[ U_{c,t12} = \text{EQ-5D-5L score at baseline of patients in the control group} \]

\[ E_{l,t0} = \text{ODI score at 12 months of patients in the intervention group} \]

\[ E_{c,t0} = \text{ODI score at baseline of patients in the control group} \]

\[ E_{l,t12} = \text{ODI score at 12 months of patients in the control group} \]

\[ E_{c,t12} = \text{ODI score at baseline of patients in the control group} \]

EO-5D-5L, Five-Level EuroQol Five-Dimensional; IC/ES, Institute for Clinical Evaluative Sciences; MOH, Ministry of Health; ODI, Oswestry Disability Index; QALY, Quality-adjusted life-years.
DISCUSSION

Identification of subgroups of patients that best respond to different interventions has long been a research priority in LBP and is an agenda driven by clinicians and researchers. Despite this, few high-quality research studies have been conducted, which are likely because of the limited understanding of important methodological criteria that are specific for these types of studies. This includes a priori designation of plausible effect modifiers, large sample sizes and a robust statistical analysis. In addition, replication of studies are often not performed, which significantly limits translation into clinical practice. This study will fill this important gap in the literature and potentially validate our promising preliminary study.

Strengths and weaknesses

A strength of this study is its design as a replication study as a critical step to test the robustness of findings of a previous preplanned analysis of effect modification in a high-quality RCT. The design of this study uses a strong RCT design with a preplanned effect modification analysis based on a specific plausible rationale founded on the results of the high-quality prior study.

Limitations of this study include the lack of blinding of patients and physiotherapists, although this is common practice in studies of exercise interventions. Further, collection of out-of-pocket costs using the Cost for Patients Questionnaire may suffer from recall bias, which is mitigated by a relatively frequent data collection. Although data on service use from administrative databases are of high quality, it will not be possible to distinguish use related to LBP. In addition, we will add a significantly large list of effect modifiers which may be prone to type I error, although we have identified the LSIQ as the primary effect modifier of interest. Finally, the results of two different Rasch analysis studies demonstrated conflicting evidence for the adequacy for calculating a summary score of the questionnaire and also demonstrated potential problems with person reliability. However, given the conflicting evidence of the two RASCH studies and strong evidence for predictive validity available, the use of this questionnaire is still appropriate. Further, there has been questions about the construct being measured by the questionnaire and whether it represented instability or another construct (eg, nociception). The additional measures of pain characteristics added into this study will allow for a better understanding of the LSIQ construct.

The results of the previously published study identified a simple and easy method to subgroup individuals into more homogenous groups of patients with LBP that led to improved efficacy of exercise. Implementation of this new approach to personalisation of exercise could be easily adopted as both treatments evaluated are currently widely used, and the method to identify subgroups of responders to each approach is straightforward, quick and at no cost. There has been an appetite from clinicians and researchers for the development of evidence-based patient subgroups, and the results of this study could result in a major advance in the field.

TRIAL STATUS

Recruitment was halted due to COVID-19. No participant has been randomised into the study.

REFERENCES

1. Koes BW, van Tulder M, Lin C-WL, et al. An updated overview of clinical guidelines for the management of non-specific low back pain in primary care. Eur Spine J 2010;19:2075–94.
2. Oliveira CB, Maher CG, Pinto RZ, et al. Clinical practice guidelines for the management of non-specific low back pain in primary care: an updated overview. Eur Spine J 2010;27:2791–803.
3. Hayden JA, van Tulder MW, Malmivaara A. Exercise therapy for treatment of non-specific low back pain. Cochrane Libr 2006;4.
null
pain patients: results of a randomized controlled trial. *Pain* 2008;138:192–207.

57 Lindström I, Ohlund C, Eek C, et al. The effect of graded activity on patients with subacute low back pain: a randomized prospective clinical study with an Operant-Conditioning behavioral approach. *Phys Ther* 1992;72:279–90.

58 Pengel LHM, Refshauge KM, Maher CG, et al. Physiotherapist-Directed exercise, advice, or both for subacute low back pain. *Ann Intern Med* 2007;146:787–96.

59 Nicholas M, Tonkin L. Persisting pain: using cognitive-behavioural principles for activity-based pain management. In: Refshauge K, Gass E, eds. *Musculoskeletal physiotherapy clinical science and evidence-based practice*. 2nd edn. Oxford: Butterworth Heinemann, 2004.

60 Anema JR, Steenstra IA, Bongers PM, et al. Multidisciplinary rehabilitation for subacute low back pain: graded activity or workplace intervention or both? *Spine* 2007;32:291–8.

61 Smeets RJEM, Vlaeyen JWS, Hidding A, et al. Instruments for clinical trials in nonspecific low back pain. *Disability and rehabilitation* 2006;7:5.

62 Nicholas M, Molloy A, Tonkin L. Manage your pain. Sydney: ABC Books, 2000.

63 Chiariotto A, Broers M, Deyo RA, et al. Core outcome measurement instruments for clinical trials in nonspecific low back pain. *Pain* 2018;159:481–95.

64 Fairbank J, Davies J, Couper J. The Oswestry low back pain disability questionnaire. *Physiotherapy* 1980;66:271–3.

65 Carles PW, Soebye LS, Stratford PF. Patients with back pain: change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. *Spine* 2008;33:90–4.

66 Vianin M. Psychometric properties and clinical usefulness of the Oswestry disability index. *J Chiropr Med* 2008;7:161–3.

67 Chiariotto A, Carles PW, Soebye LS, et al. A systematic review highlights the need to investigate the content validity of patient-reported outcome measures for physical functioning in patients with low back pain. *J Clin Epidemiol* 2018;95:73–93.

68 Pengel LHM, Refshauge KM, Maher CG. Responsiveness of pain, disability, and physical impairment outcomes in patients with low back pain. *Spine* 2004;29:879–83.

69 Hall AM, Maher CG, Latimer J, et al. The patient-specific functional scale is more responsive than the Roland Morris disability questionnaire when activity limitation is low. *Eur Spine J* 2011;20:79–86.

70 Hayden J, van Tulder MW, Malmivaara A, et al. Exercise therapy for treatment of non-specific low back pain. *Cochrane Database Syst Rev* 2005;35.

71 Devlin NJ, Brooks R. EQ-5D and the EuroQol group: past, present and future. *Appl Health Econ Health Policy* 2017;15:127–37.

72 Xie F, Pullenayagum E, Gaebel K, et al. A time trade-off-derived value set of the EQ-5D-5L for Canada. *Med Care* 2016;54:98–105.

73 CADTH. Guidelines for the economic evaluation of health technologies. Ottawa: CADTH, 2017.

74 Lin C-WC, Haas M, Maher CG, et al. Cost-effectiveness of general practice care for low back pain: a systematic review. *Eur Spine J* 2011;20:1012–23.

75 Deyo RA, Dworkin SF, Arnertt D, et al. Report of the NIH Task force on research standards for chronic low back pain. *The Journal of Pain* 2014;15:569–85.

76 Dworkin RH, Turk DC, Farrar JT, et al. Core outcome measures for chronic pain clinical trials: IJM PACT recommendations. *Pain* 2005;113:9–19.

77 Araujo AC, da Cunha Menezes Costa L, de Oliveira CBS, et al. Measurement properties of the Brazilian-Portuguese version of the lumbar spine instability questionnaire. *Spine* 2017;42:E810–4.

78 Saragiotto BT, Maher CG, New CH, et al. Clinimetric testing of the lumbar spine instability questionnaire. *J Orthop Sports Phys Ther* 2018;48:315–22.

79 Kendall N, Linton S, Main C. Guide to assessing psychosocial yellow flags in acute low back pain: risk factors for long term disability and work loss. Accident rehabilitation and compensation: a Cochrane synthesis of New Zealand and the. Wellington: National Health Committee, 1997.

80 Fuhr FF, Fagundes FR, Manzoni AC. Orosb musculoskeletal pain screening questionnaire short-form and start back screening tool: correlation and agreement analysis. *Spine* 2016;41:E931–8.

81 Langenfeld A, Bastiaansen C, Brunner F, et al. Validation of the Orosb musculoskeletal pain screening questionnaire in patients with chronic neck pain. *BMC Res Notes* 2018;11:161.

82 Opsomer E, Hilfinger R, Raval-Roland B, et al. Test-retest reliability of the Orosb musculoskeletal pain screening questionnaire and the situational pain scale in patients with chronic low back pain. *Swiss Med Wkly* 2013;143:w13903.

83 Plieoe E, Neubauer E, Pfeifer AC, et al. Predicting persistent back symptoms by psychosocial risk factors: validity criteria for the OMPSQ and the HKF-R 10 in Germany. *PloS One* 2016;11:e0158850.

84 Ruokolainen O, Haapea M, Linton S, et al. Construct validity and reliability of Finnish version of Orosb musculoskeletal pain screening questionnaire. *Scand J Pain* 2016;13:148–53.

85 Nebbett R, Hartzell MM, Mayer TG, et al. Establishing clinically meaningful severity levels for the Tampa scale for Kinesiophobia (TSP-13). *Eur J Pain* 2016;20:701–6.

86 Lawson K, Reesor KA, Keele F, et al. Dimensions of pain-related cognitive coping: cross-validation of the factor structure of the coping strategy questionnaire. *Pain* 1990;43:195–204.

87 Robinson ME, Riley JL, Myers CD, et al. The coping strategies questionnaire: a large sample, item level factor analysis. *Clin J Pain* 1997;15:43–9.

88 Riddle DL, Jensen MP. Construct and criterion-related validity of brief pain coping scales in persons with chronic knee osteoarthritis pain. *Pain Med* 2013;14:265–75.

89 Serra ML, Agost F, et al. Translation, cross-cultural adaptation, reliability, and validity of the German version of the coping strategies questionnaire (CSO-D). *J Pain* 2006;7:327–36.

90 Freynhagen R, Baron R, Gockel U, et al. pain DETECT: a new screening questionnaire to identify neuropathic components in chronic pain. *Curr Med Res Opin* 2006;22:1911–20.

91 Ramasamy A, Martin ML, Blum SI, et al. Assessment of patient-reported outcome instruments to assess chronic low back pain. *Pain Med* 2018;17:1098–110.

92 Keller T, Freynhagen R, Tolle TR, et al. A retrospective analysis of the long-term test-retest stability of pain descriptors of the painDETECT questionnaire. *Curr Med Res Opin* 2016;32:343–9.

93 Smart KM, Blake C, Staines A, et al. The Discriminative Validity of “Nociceptive,” “Peripheral Neuropathic,” and “Central Sensitization” as Mechanisms-based Classifications of Musculoskeletal Pain. *Clin J Pain* 2011;27:65–83.

94 Rabey M, Kendell M, Godden C, et al. Start back tool risk stratification is associated with changes in movement profile and sensory discrimination in low back pain: a study of 290 patients. *Eur J Pain* 2019;23:823–34.

95 Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. *Appl Psychol Measur* 1977;1:385–401.

96 Sullivan MJL, Bishop SR, Pivik J. The pain catastrophizing scale: development and validation. *Psychol Assess* 1995;7:524–32.

97 Neblett R, Mayer TG. The central sensitization inventory (CsI): some gender as predictor and cognitive coping: cross-measurement of the lumbar spine instability questionnaire. *Eur J Pain* 2013;143:W13903.

98 Rabef M, Kendell M, Godden C, et al. Start back tool risk stratification is associated with changes in movement profile and sensory discrimination in low back pain: a study of 290 patients. *Eur J Pain* 2019;23:823–34.

99 Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. *Appl Psychol Measur* 1977;1:385–401.

100 Sullivan MJL, Bishop SR, Pivik J. The pain catastrophizing scale: development and validation. *Psychol Assess* 1995;7:524–32.

101 Neblett R, Mayer TG. The central sensitization inventory (CsI): some gender as predictor and cognitive coping: cross-measurement of the lumbar spine instability questionnaire. *Eur J Pain* 2013;143:W13903.

102 Rabef M, Kendell M, Godden C, et al. Start back tool risk stratification is associated with changes in movement profile and sensory discrimination in low back pain: a study of 290 patients. *Eur J Pain* 2019;23:823–34.

103 Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. *Appl Psychol Measur* 1977;1:385–401.

104 Sullivan MJL, Bishop SR, Pivik J. The pain catastrophizing scale: development and validation. *Psychol Assess* 1995;7:524–32.