Exploring the heterogeneity of musculoskeletal pain

Lingxiao Chen

A thesis submitted in fulfilment of the requirement for the degree of Doctor of Philosophy

Faculty of Medicine and Health

The University of Sydney

July 2021

Supervisors' Statement

As supervisors of Lingxiao Chen's doctoral work, we certify that we consider his thesis "Exploring the heterogeneity of musculoskeletal pain" sufficiently well presented to be examined.

Primary Supervisor: Professor Manuela L Ferreira

Institute of Bone and Joint Research, Faculty of Medicine and Health

The University of Sydney

Date: 21 July 2021

Auxiliary Supervisor: Professor Nigel K Arden

Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences

The University of Oxford

Date: 21 July 2021

Auxiliary Supervisor: Dr Paula R Beckenkamp

School of Health Sciences, Faculty of Medicine and Health

The University of Sydney

Date: 21 July 2021

Auxiliary Supervisor: Professor Paulo R Ferreira

School of Health Sciences, Faculty of Medicine and Health

The University of Sydney

Date: 21 July 2021

Candidates Statement

I, Lingxiao Chen, hereby declare that this information is my own work and that it contains no material previously published or written by another person except where acknowledged in the text. Nor does it contain material which has been accepted for the award of another degree.

I, Lingxiao Chen, understand that if I am awarded a higher degree for my thesis entitled 'Exploring the heterogeneity of musculoskeletal pain' being lodged herewith for examination, the thesis will be lodged in the University library and be available immediately for use. I agree that the University Librarian (or in the case of a department, the Head of Department) may supply a photocopy of microform of the thesis to an individual for research or study or to a library.

Date: 21 July 2021

Table of Contents

Acknowledgements

Firstly, I want to thank my family. My mother Mrs Ainong Zhang and my father Mr Bangyan Chen supported me in every corner. My girlfriend Dr Yujie Chen has always been supportive, especially when I felt anxiety, although she is also busy with the PhD.

Secondly, I want to thank my supervisors. My primary supervisor Prof Manuela Ferreira gave the me the freedom to explore the area I am interested in. With her careful guidance, I made tremendous progresses during my PhD journey. My auxiliary supervisor Professor Nigel Arden was the person who led me to the area of epidemiology and taught me how to build an excellent research question. My auxiliary supervisor Dr Paula Beckenkamp always supported me. My auxiliary supervisor Prof Paulo Ferreira gave me the opportunity to lead an external-funded project and created a lot of collaborative opportunities.

Thirdly, I want to thank my collaborators, who I was fortunate enough to work with, including (but not limited to): Dr Eduardo Caputo, Prof Shiqing Feng, Dr Ralph Stanford, Dr Munkh-erdene Bayartai, Dr David Anderson, Dr Romain Perera, Dr Maja Radojčić, Dr Deborah Hart, Prof Tim Spector, Prof Natasha Nassar, Prof David Preen, Prof John Hopper, Dr Shuai Li, Dr Minh Bui, Dr Xia Wang, Ms Baoyi Shi, Ms Emma Ho, and Ms Carolina Fritsch.

Finally, I want to thank all others who helped me, including (but not limited to):

Prof David Hunter, Prof Rob Herbert, Prof Jim Elliott (my mentor), Prof Emmanuel Stamatakis, Prof Frank Harrell, Prof Sander Greenland, Prof Wei Zeng, Prof Hengxing Zhou, Prof Tiansong Zhang, A/Prof Maarten van Smeden, A/Prof Ellie Murray, Dr Kai Fu, Dr Zhirui Zhou, Dr Wei Wang, and Ms Sai Zhao.

Published or Submitted Papers

Chen L., Radojčić M.R., Perera R.S., Beckenkamp P.R., Ferreira P.H., Hart D.J., Spector T.D., Arden N.K., Ferreira M.L. Association of analgesic use trajectories with mortality and quality of life in middle-aged, community-dwelling women: a population-based cohort study. *Under review (Pain)*

Chen L., Ferreira P.H., Stanford R., Beckenkamp P.R., Bayartai M., Anderson D.B., Arden N.K., Ferreira M.L**.** Effectiveness and Safety of Surgical, Invasive Treatments and Conservative Care for Degenerative Lumbar Spinal Stenosis: A Systematic Review and Network Meta-analysis. *Under review (Physical Therapy)*

Chen L., Ferreira M.L., Nassar N., Preen D., Hopper J.L., Li S., Bui M., Beckenkamp P.R., Shi B., Arden N.K., Ferreira P.R. Association of Chronic Musculoskeletal Pain with Mortality Among Middle-Aged UK Participants: A Population-Based Cohort Study with Mediation Analysis. EClinicalMedicine. 2021 Nov 14;42:101202.

Chen L., Perera R.S., Radojčić M.R., Beckenkamp P.R., Ferreira P.H., Hart D.J., Spector T.D., Arden N.K., Ferreira M.L. Association of Lumbar Spine Radiographic Changes With Severity of Back Pain-Related Disability Among Middle-aged, Community-Dwelling Women. JAMA Netw Open. 2021;4(5):e2110715.

Chen L., Ferreira ML, Beckenkamp PR, Caputo EL, Feng S, Ferreira PH. Comparative Efficacy and Safety of Conservative Care for Pregnancy-Related Low Back Pain: A Systematic Review and Network Meta-analysis. Phys Ther. 2021 Feb 4;101(2):pzaa200.

Chen L., H Ferreira P, R Beckenkamp P, L Ferreira M. Comparative efficacy and safety of surgical and invasive treatments for adults with degenerative lumbar spinal stenosis: protocol for a network meta-analysis and systematic review. BMJ Open. 2019 Apr 4;9(4):e024752.

Presentations

L Chen, M Ferreira, P Beckenkamp, E Caputo, S Feng, P Ferreira. Comparative efficacy and safety of conservative care for pregnancy-related low back pain: a systematic review and network meta-analysis. Spine Symposium. September 2019. Sydney, Australia. Poster Presentation.

L Chen, M Ferreira, P Beckenkamp, E Caputo, P Ferreira. Comparative efficacy and safety of conservative care for pregnancy-related low back pain: a systematic review and network meta-analysis. *Osteoarthritis Research Society International*. May 2019. Toronto, Canada. Poster Presentation.

Preface

This thesis is arranged in eight chapters, with an introduction (**Chapter One**), six submitted or published papers (**Chapter Two** to **Chapter Seven**), and a conclusion (**Chapter Eight**). The University of Sydney allows published papers that arose during the candidature to be included in the thesis.

Chapter One is an introduction to the thesis and provides the background on musculoskeletal pain and main statistical methods used. **Chapter Two** is a systematic review and network meta-analysis evaluating the effectiveness and safety of conservative care approaches for pregnancy-related back pain. This chapter is presented as published in *Physical Therapy*. **Chapter Three** is a research protocol of a network meta-analysis and systematic review assessing the effectiveness and safety of surgical and invasive procedures for adults with degenerative lumbar spinal stenosis. This chapter is presented as published in *BMJ Open*. **Chapter Four** presents the full systematic review and network meta-analysis on the comparative effectiveness and safety of surgery, invasive treatments and conservative care for people with degenerative lumbar spinal stenosis. This chapter is presented as submitted for publication in *Physical Therapy*. **Chapter Five** is a cohort study examining both cross-sectional and longitudinal associations between lumbar radiographic changes and the severity of back pain-related disability. This chapter is presented as published in *JAMA Network Open*. **Chapter Six** is a cohort study conducted to identify distinct trajectories of analgesic use and the potential association of these trajectories with mortality and quality of life. This chapter is presented as submitted for publication in *Pain*. **Chapter Seven** is a cohort study quantifying the association between chronic musculoskeletal pain and all-cause mortality and the extent to which this association is mediated by physical activity, smoking status, alcohol consumption, and opioid use. This chapter is presented as published in *EClinicalMedicine*. **Chapter Eight** is an overview of the thesis and discusses the clinical implications of the findings and recommendations for future research.

Each chapter contains its own reference list. Ethical approval was obtained from the Waltham Forest and Redbridge Local Research Ethics Committee, UK for the studies presented in **Chapter**

Five and **Chapter Six**, and the North West Multi-centre Ethics Committee, UK, for the study presented in **Chapter Seven**. The remaining chapters did not require ethical approval.

Abstract

Musculoskeletal pain often includes pain in the back, neck, knee, and hip, and is associated with a substantial financial and personal burden. Eight chapters are included in this thesis that aims to improve the understanding of the heterogeneity in treatment effects and prognosis of musculoskeletal pain. Four issues were identified: i) people with different pain phenotypes (i.e. back pain with or without neurological deficit) or with distinct underlying health conditions (e.g. pregnancy-related back pain) may respond differently to treatment strategies; ii) people with chronic back pain and presenting different radiological phenotypes may experience different course of the disease; iii) different patterns of analgesic use over time may be associated with different long term health status; iv) different types and number of sites of musculoskeletal pain may be associated with different clinical prognoses. **Chapter One** is an introduction to musculoskeletal pain.

Pregnancy is a special time during a woman's life. Research shows, however, that more than one-half of pregnant women experience pregnancy-related back pain. Despite many conservative strategies being commonly recommended for pregnancy-related back pain, little is known about the comparative effectiveness and safety of these approaches. Thus, a systematic review and network meta-analysis was conducted and is presented in **Chapter Two** comparing the effectiveness and safety of different types of conservative care for pregnancy-related back pain. The results showed that, for women with back pain during pregnancy, progressive muscle relaxation therapy (mean difference [MD]: −3.96, 95% confidence interval [CI]: −7.19 to −0.74; moderate-quality evidence) and Kinesio Taping (MD: −3.71, 95% CI: −6.55 to −0.87; low-quality evidence) resulted in small reductions in pain intensity (Visual Analog Scale, range = 0 to 10) compared with placebo. Moderate-quality evidence suggested that transcutaneous electrical nerve stimulation results in a moderate improvement in function (MD: −6.33, 95% CI: −10.61 to −2.05; Roland Morris Disability Questionnaire, range = 0–24) compared with placebo. This study bridges an important gap in the evidence regarding optimal management of pregnancy-related back pain.

Degenerative lumbar spinal stenosis affects 11% of the population and is often managed with surgery or other invasive treatment options (e.g., epidural injection). Similar to pregnancy-related back pain, little is known about the comparative effectiveness and safety of available management approaches for lumbar spinal stenosis. Thus, a protocol for a systematic review and network meta-analysis is presented in **Chapter Three** and a network meta-analysis is presented in **Chapter Four** comparing the effectiveness and safety of surgery, invasive treatments, and conservative care for degenerative lumbar spinal stenosis. Overall, the results showed no statistically significant differences between conservative approaches and any surgical or invasive interventions in the primary outcomes of physical function and all-cause mortality. Likewise, no significant differences between groups were observed for back pain, mobility, or treatment withdrawal due to any reason. Although interspinous device (MD: -2.05, 95% CI: -3.98 to -0.12), midline splitting decompression (MD: -2.47, 95% CI: -4.45 to -0.5) and conventional open decompression (MD: -1.80, 95% CI: -3.49 to -0.11) were statistically superior to conservative care for short-term leg pain (0-10-point Visual Analog Scale), the difference was clinically unimportant. Conservative care was associated with lower odds of intervention-related adverse events when compared with other surgical interventions, except for endoscopic decompression. When comparing across all surgical interventions, interspinous device was the procedure associated with the highest odds of reoperation. This network meta-analysis provides the reader with the most comprehensive understanding of the current evidence for treatments of degenerative lumbar spinal stenosis.

Current international practice guidelines do not recommend routinely using diagnostic imaging in the management of back pain. However, spinal diagnostic imaging is still widely used in clinical practice. Previous studies, using mostly cross-sectional data, provide conflicting evidence of a potential association between lumbar spine radiographic changes and the severity of back pain–related disability. Such conflicting evidence may be associated with the wide use of unnecessary diagnostic imaging of the lumbar spine. **Chapter Five** aimed to examine both cross-sectional and longitudinal associations between lumbar spine radiographic changes and the severity of back pain-related disability among middle-aged, community-dwelling women. The study found that, among 650 women (mean [SD] age, 61.3 [5.9] years) whose data were included in the cross-sectional analyses, and 443 women (mean [SD] age, 60.6 [6.0] years) included in the longitudinal

analyses, there was no evidence to support an association between higher number of lumbar segments with radiographic changes (K-L grade, osteophytes, and disc space narrowing) and more severe back pain–related disability (e.g., cross-sectional analyses using the K-L grade; 1 segment vs 0 segment: adjusted odds ratio, 1.22, 95% CI, 0.76 to 1.96]). No interactions were found of an association between lumbar spine radiographic changes and the severity of back pain–specific disability with age, body mass index, or smoking status. The findings suggest that the changes detected in lumbar radiographs provide limited value for decision-making regarding the management of back pain in this population.

Analgesics are widely used and at a steadily increasing rate over the past years. Given the many associated adverse events of analgesics and the potential impact their long-term use may have on the patient's general health, it is relevant to better understand whether different patterns of analgesic use may influence adverse outcomes. **Chapter six** aimed to identify distinct trajectories of analgesic use and the potential association of these trajectories with mortality and quality of life. Among 804 women, three distinct trajectories of analgesic use were identified: (i) 'no use' group (691, 85.9%); (ii) 'increasing probability to use' group (73, 9.1%); and (iii) 'constant analgesic use' group (40, 5.0%). Compared with the 'no use' group, the 'constant analgesic use' group was associated with 2.15 times higher risk of all-cause mortality (95% CI: 1.18 to 3.91). No association between cause-specific mortality and pattern of analgesic use was found. Worse quality of life in terms of physical function, role limitations due to physical health and pain was associated with constant and high probability of using analgesics, and with increased probability of using analgesics. This study showed that, in this cohort of middle-aged women, a small group of women presented a high and constant probability of using analgesic over the study period and a markedly higher risk of all-cause mortality compared to those with no or low probability of using analgesics. The findings indicate the need for public health initiatives addressing the potential drivers of a high and constant probability of using analgesics, including better communication between physicians and patients and effective education for the community-dwelling people.

Chronic musculoskeletal pain is common, and although it has been linked with increased risk of mortality, the nature of this association is still uncertain. **Chapter seven** aimed to quantify the association between chronic musculoskeletal pain and all-cause mortality and to investigate the extent to which this association was mediated by physical activity, regular opioid use, current smoking, and regular alcohol consumption. Of the 384,367 included participants, 54.2% were women, with mean (SD) age of 57 (8) years. A total of 187,269 participants reported chronic musculoskeletal pain. Neck or shoulder pain only, back pain only and hip pain only were associated with higher risk of all-cause mortality. Higher number of pain sites was associated with increased risk of all-cause mortality compared to having no pain (e.g., four sites vs no site of pain, Hazzard Ratio [HR] 1.46, 95% CI 1.35 to 1.57). The single mediator analyses showed the following mediating proportions of the association between chronic musculoskeletal pain and all-cause mortality: 8.0% to 15.7% for physical activity; 32.5% to 79.0% for opioid use; 14.6% to 29.8% for smoking status and 2.4% to 17.5% for alcohol consumption. The multiple mediator analyses showed that the mediating proportions for all four mediators ranged from 53.4% to 122.6%. Among participants with one pain site, chronic musculoskeletal pain was not associated with all-cause mortality. Among participants with two or more pain sites, however, the effect estimate reduced substantially, e.g., for two pain sites, HR reduced from 1.25 (95% CI: 1.21 to 1.30) to 1.07 (95% CI: 1.01 to 1.11). The results suggest that supporting healthy lifestyle behaviour as well as opioids deprescription is an important strategy to decrease the mortality risk associated with chronic musculoskeletal pain.

CHAPTER ONE

Introduction

Introduction

Background

The global burden of musculoskeletal pain is substantial, especially in the older population. Based on the results from the 2019 Global Burden of Disease, the global prevalence (all ages and both sexes) for rheumatoid arthritis, osteoarthritis, low back pain, neck pain and gout is 18.6 million (0.25%), 0.5 billion (7.1%), 0.6 billion (7.6%), 0.2 billion (3.0%) and 53.9 million (0.72%), respectively[1]. The prevalence rate is higher for those who are 65 years of age or older: 1.7% for rheumatoid arthritis, 64.8% for osteoarthritis, 38.5% for low back pain, 12.1% for neck pain and 6.4% for gout[1]. These conditions are also leading causes of disability worldwide. For example, low back pain has been considered the number one cause of years lived with disability from 1990 to 2017, with neck pain ranking the number ten[2].

There is substantial heterogeneity in the clinical course of these conditions as well as in how people respond to different treatment options used to manage musculoskeletal pain[3-14]. I have identified the following issues which may be associated with these levels of heterogeneity: i) people with different pain phenotypes (i.e. back pain with or without neurological deficit) or with distinct underlying health conditions (e.g. pregnancy-related back pain) may respond differently to treatment strategies; ii) people with chronic back pain and presenting different radiological phenotypes may experience different course of the disease; iii) different patterns of analgesic use over time may be associated with different long term health status; iv) different types and number of sites of musculoskeletal pain may be associated with different clinical prognoses.

Treatment for pregnancy-related back pain

Pregnancy is a special time during a woman's life, however more than one-half of pregnant women experience pregnancy-related back pain[15]. About 50% of women with pregnancy-related back pain present with persistent symptoms 12 months postpartum, and 20% remain with symptoms three years postpartum[16,17]. Pregnancy-related pelvic girdle pain (PGP) is another type of musculoskeletal pain, which presents different prognosis compared with pregnancy-related back pain (women with pregnancy-related PGP are more disabled than pregnancy-related back pain)[18].

The management of pregnancy-related PGP and back pain is largely focussed on conservative approaches to relieve pain and improve function during pregnancy, although very few clinical practice guidelines offer specific recommendations for these conditions. For instance, one of the few guidelines to provide specific recommendations, the Irish clinical practice guideline for PGP management in pregnancy and post-partum endorses individualised physiotherapy tailored programs, exercise, and education for back pain in pregnancy and post-partum[19]. The guideline also mentions that manipulation or mobilisation of pelvic joints and acupuncture may be used; whilst massage and pelvic belts should not be used as a single intervention; paracetamol, codeine-based preparations and non-steroidal anti-inflammatory drugs should be used cautiously.

Despite many conservative strategies being commonly recommended for pregnancy-related back pain, little is known about the comparative effectiveness and safety among these approaches as these recommendations are based on the results of two-treatment comparisons reported in existing trials or systematic reviews. Thus, a network meta-analysis and systematic review of randomised controlled trials and observational studies was conducted and is presented in **Chapter Two** evaluating the comparative effectiveness and safety among numerous conservative care approaches for pregnancy-related back pain.

Treatment for degenerative lumbar spinal stenosis

Degenerative lumbar spinal stenosis is a specific type of back pain that affects 11% of the population around the world and is recognised as one of the most debilitating forms of back pain [20,21]. Usually, this condition is the result of a decreased spinal canal diameter due to structural changes that occur with age[22], with patients typically experiencing pain, numbness and/or fatigue in one or both lower limbs that is worsened during walking and standing and alleviated with forward bending or sitting[8] – also described as neurogenic claudication. Occasionally patients may present with spinal instability or degenerative spondylolisthesis, which can change the way their condition is managed or how they respond to some treatment approaches[21]. The degree of stenosis of the spinal canal also varies substantially among patients, and there is little correlation between the degree of stenosis and the severity or location (i.e., unilateral vs bilateral) symptoms[21]. These variations in patient phenotype challenge the choice of treatment options[21].

International clinical practice guidelines will usually recommend surgical interventions (e.g., decompression, fusion, and interspinous spacer), injections (e.g., epidural injection and adhesiolysis), and nonsurgical interventions (e.g., physical therapy and bracing)[23-33]. However, similar to pregnancy-related back pain, little is known about the comparative effectiveness and safety of these available approaches, i.e., surgical interventions and conservative care and most guideline recommendations are based on head-to-head comparisons between two interventions. Thus, a research protocol is outlined in **Chapter Three** and the formal systematic review and network meta-analysis is presented in **Chapter Four** assessing the effectiveness and safety of surgical, invasive treatments and conservative care for people with degenerative lumbar spinal stenosis.

Imaging for back pain

A common source of clinical heterogeneity among people with back pain is the large range of radiological phenotypes observed in this population. Patients with back pain may present with degenerative changes to the intervertebral spaces, spondylolisthesis, or osteoporotic fractures. Research, however, has not been able to fully establish a reliable correlation between symptom severity or clinical prognosis and radiological findings among people with back pain, as highlighted below. For that reason, routine diagnostic imaging for patients with back pain is in general discouraged. For example, the clinical practice guidelines developed by the American College of Physicians[34] highlights only three indications for diagnostic imaging:

- Immediate imaging: major risk factors for cancer, risk factors for spinal infection, risk factors for or signs of the cauda equina syndrome, severe or progressive neurologic deficits.
- Imaging after a trial of therapy: minor risk factors for cancer, risk factors for inflammatory back disease or vertebral compression fracture, signs or symptoms of radiculopathy, risk factors for or symptoms of symptomatic spinal stenosis.
- Repeated imaging: new or changed low back symptoms.

Similar recommendations are seen in the National Institute for Health and Care Excellence[35], which recommends that imaging should be only considered in specialist settings of care (for example, a musculoskeletal interface clinic or hospital) if the results are likely to change the management. However, diagnostic imaging is still widely used. A recent systematic review and meta-analysis, which included 45 studies with 4 million imaging requests/events, showed that[36]:

- Simple imaging (defined as plain radiography or ultrasound): no significant change in the proportion of requests over 20 years with 21.2% of patients with back pain (clinical setting includes emergency care and primary care) receiving a referral for imaging in 1995 and 21.3% in 2015.
- Complex imaging (defined as computed tomography, magnetic resonance imaging or nuclear bone scan): a significant increase over 20 years with 7.4% in 1995 and 11.4% in 2015.

Previous studies, using mostly cross-sectional data, provided conflicting evidence of any association between lumbar spine radiographic changes and the severity of back pain–related disability[37,38]. Such conflicting evidence may be associated with widely unnecessary diagnostic imaging of the lumbar spine. Thus, a cohort study was conducted and is presented in **Chapter Five** examining both cross-sectional and longitudinal associations between lumbar radiographic changes and the severity of back pain–related disability using composite scores that combined the number of segments and type of changes in terms of K-L grade, disc space narrowing, and osteophytes.

Analgesic use

Pharmacologic treatments including opioid, nonsteroidal anti-inflammatory drugs (NSAIDs) and acetaminophen, are mainstream options to alleviate musculoskeletal pain, although the effects of these treatments on pain and function are small to moderate[39,40]. In general, analgesics could be classified as opioid analgesics (e.g., tramadol and morphine), non-opioid analgesics (e.g., paracetamol, aspirin, and ibuprofen), and compound (e.g., Co-Codamol - codeine and paracetamol and Co-Dydramol – dihydrocodeine and paracetamol). On 07 April 2021, the National Institute

for Health and Care Excellence (NICE) released a new guideline titled *Chronic pain (primary and secondary) in over 16s: assessment of all chronic pain and management of chronic primary pain*[41]. The NICE guidelines mentioned that these analgesics should not be used as first line care of chronic primary pain (including musculoskeletal pain), defined based on the following International Classification of Diseases – 11 (ICD-11) criteria[42]:

- pain that persists or recurs for longer than 3 months, AND
- pain that is associated with significant emotional distress (e.g., anxiety, anger, frustration, or depressed mood) and/or significant functional disability (interference in activities of daily life and participation in social roles), AND
- the symptoms are not better accounted for by another diagnosis.

However, although current guidelines refrain from recommending the prescription of analgesics, especially opioids, for the management of chronic pain, these are still widely used by patients with musculoskeletal pain. Different patterns of analgesic use in terms of type of medication and duration of use can be identified in clinical practice. These differences might be associated with distinct long term health status. Thus, a cohort study presented in **Chapter Six** was conducted to identify distinct trajectories of analgesic use and identify the association of these trajectories with mortality and quality of life.

Chronic musculoskeletal pain and mortality

The nature and magnitude of the association between chronic musculoskeletal pain and the risk of mortality is still unclear. There is substantial heterogeneity in the definitions used for chronic musculoskeletal pain among previous studies, which might explain the lack of clarity in this field. For example, the suggested standard definition of 'chronic pain' is 'pain duration of at least 3 months'[42]. However, existing studies in this field have failed to use this definition. Holmberg et al defined chronic as pain in the last 14 days, whereas Roseen defined chronic as pain symptoms reported in the past 12 months "most of the time" or "constantly", both at baseline and first follow-up visit[43,44]. A second issue is the heterogeneity in the design of previous studies, including the selection of the control group. Previous studies have compared the risk of mortality in people with

musculoskeletal pain with that among those with either other types of chronic pain, or with acute or subacute musculoskeletal pain. From a clinical and epidemiological perspective, the control group should be defined as the participants without pain to better inform practice and policy. One final issue is related to how chronic musculoskeletal pain may increase the risk of mortality. For instance, the role of opioid use as a potential mediator between chronic musculoskeletal pain and risk of mortality has not been well established[44,45]. Moreover, whilst a previous study explored the role of three lifestyle behaviours (physical activity, smoking status, and alcohol consumption) individually[45], the combined role of lifestyle behaviour, i.e., the simultaneous influence of these three factors is still unknown. These results would have a substantial impact on policy and the development of preventative strategies. Thus, a cohort study was performed and is presented in **Chapter Seven** quantifying the association between chronic musculoskeletal pain and all-cause mortality and investigating the extent to which this association is mediated by physical activity, smoking status, alcohol consumption, and opioid use.

Challenges of assessment

Most of the research included in this thesis is based on patient reported outcomes (e.g., pain, physical function, and quality of life). There is some uncertainty associated with self-reported outcomes, which may affect the results and needs to be acknowledged. For example, pain intensity can be measured in a variety of ways: pain at its worst, pain over the last 24 hours, average pain in the last 7 days. Additionally, diagnoses based on spinal imaging might be subject to error due to low interrater reliability.

Aims of the thesis

The aims of this thesis were:

- To evaluate the effectiveness and safety of conservative care approaches for pregnancy-related back pain. (**Chapter Two**)
- To develop a research protocol (**Chapter Three**) and conduct a full network meta-analysis (**Chapter Four**) to assess the effectiveness and safety of surgical, invasive treatments and conservative care for people with degenerative lumbar spinal stenosis.

- To examine both cross-sectional and longitudinal associations between lumbar radiographic changes and the severity of back pain-related disability. (**Chapter Five**)
- To identify distinct trajectories of analgesic use and identify the association of these trajectories with mortality and quality of life. (**Chapter Six**)
- To quantify the association between chronic musculoskeletal pain and all-cause mortality, and to investigate the extent to which this association is mediated by physical activity, smoking status, alcohol consumption, and opioid use. (**Chapter Seven**)

References

1. 2019 GBD Results Tool [Available from: http://ghdx.healthdata.org/gbd-results-tool accessed May 15 2021.

2. James SL, Abate D, Abate KH, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet.* 2018;392(10159):1789-1858.

3. Burmester GR, Pope JE. Novel treatment strategies in rheumatoid arthritis. *The Lancet.* 2017;389(10086):2338-2348.

4. Cohen SP, Hooten WM. Advances in the diagnosis and management of neck pain. *Bmj.* 2017;358.

5. McInnes IB, Schett G. Pathogenetic insights from the treatment of rheumatoid arthritis. *The Lancet.* 2017;389(10086):2328-2337.

6. Aletaha D, Smolen JS. Diagnosis and management of rheumatoid arthritis: a review. *Jama.* 2018;320(13):1360-1372.

7. Foster NE, Anema JR, Cherkin D, et al. Prevention and treatment of low back pain: evidence, challenges, and promising directions. *The Lancet.* 2018;391(10137):2368-2383.

8. Hartvigsen J, Hancock MJ, Kongsted A, et al. What low back pain is and why we need to pay attention. *The Lancet.* 2018;391(10137):2356-2367.

9. Dehlin M, Jacobsson L, Roddy E. Global epidemiology of gout: prevalence, incidence, treatment patterns and risk factors. *Nature Reviews Rheumatology.* 2020;16(7):380-390.

10. Hansildaar R, Vedder D, Baniaamam M, Tausche A-K, Gerritsen M, Nurmohamed MT. Cardiovascular risk in inflammatory arthritis: rheumatoid arthritis and gout. *The Lancet Rheumatology.* 2020.

11. Safiri S, Kolahi A-A, Hoy D, et al. Global, regional, and national burden of neck pain in the general population, 1990-2017: systematic analysis of the global burden of disease study 2017. *Bmj.* 2020;368.

12. Safiri S, Kolahi A-A, Smith E, et al. Global, regional and national burden of osteoarthritis 1990-2017: a systematic analysis of the Global Burden of Disease Study 2017. *Annals of the rheumatic diseases.* 2020;79(6):819-828.

13. Xia Y, Wu Q, Wang H, et al. Global, regional and national burden of gout, 1990–2017: a systematic analysis of the Global Burden of Disease Study. *Rheumatology.* 2020;59(7):1529-1538.

14. Katz JN, Arant KR, Loeser RF. Diagnosis and treatment of hip and knee osteoarthritis: a review. *Jama.* 2021;325(6):568-578.

15. Sencan S, Ozcan-Eksi EE, Cuce I, Guzel S, Erdem B. Pregnancy-related low back pain in women in Turkey: prevalence and risk factors. *Annals of physical and rehabilitation medicine.* 2018;61(1):33-37.

16. Norén L, Östgaard S, Johansson G, Östgaard HC. Lumbar back and posterior pelvic pain during pregnancy: a 3-year follow-up. *European spine journal.* 2002;11(3):267-271.

17. Padua L, Caliandro P, Aprile I, et al. Back pain in pregnancy: 1-year follow-up of untreated cases. *European Spine Journal.* 2005;14(2):151-154.

18. Vermani E, Mittal R, Weeks A. Pelvic girdle pain and low back pain in pregnancy: a review. *Pain Practice.* 2010;10(1):60-71.

19. Clinical practice guideline: management of pelvic girdle pain in pregnancy and post-partum. Chartered physiotherapists Women's health and continence and Directorate of Strategy and Clinical Programmes Health Service Executive. 2014.

https://www.hse.ie/eng/services/publications/clinical-strategy-and-programmes/management-of-pelvic-girdle-pain-in-pregnancy-and-post-partum.pdf.

20. Jensen RK, Jensen TS, Koes B, Hartvigsen J. Prevalence of lumbar spinal stenosis in general and clinical populations: a systematic review and meta-analysis. *European Spine Journal.* 2020:1-21.

21. Jensen RK, Harhangi BS, Huygen F, Koes B. Lumbar spinal stenosis. *bmj.* 2021;373.

22. Chen L, Ferreira PH, Beckenkamp PR, Ferreira ML. Comparative efficacy and safety of surgical and invasive treatments for adults with degenerative lumbar spinal stenosis: protocol for a network meta-analysis and systematic review. *BMJ open.* 2019;9(4):e024752.

23. Boswell MV, Shah RV, Everett CR, et al. Interventional techniques in the management of chronic spinal pain: evidence-based practice guidelines. *Pain physician.* 2005;8(1):1-47.

24. Chou R, Qaseem A, Snow V, et al. Diagnosis and treatment of low back pain: a joint clinical practice guideline from the American College of Physicians and the American Pain Society. *Annals of internal medicine.* 2007;147(7):478-491.

25. Watters III WC, Baisden J, Gilbert TJ, et al. Degenerative lumbar spinal stenosis: an evidence-based clinical guideline for the diagnosis and treatment of degenerative lumbar spinal stenosis. *The spine journal.* 2008;8(2):305-310.

26. Chou R, Loeser JD, Owens DK, et al. Interventional therapies, surgery, and interdisciplinary rehabilitation for low back pain: an evidence-based clinical practice guideline from the American Pain Society. *Spine.* 2009;34(10):1066-1077.

27. Manchikanti L, Boswell MV, Singh V, et al. Comprehensive evidence-based guidelines for interventional techniques in the management of chronic spinal pain. *Pain Physician.* 2009;12(4):699-802.

28. Kreiner DS, Shaffer WO, Baisden JL, et al. An evidence-based clinical guideline for the diagnosis and treatment of degenerative lumbar spinal stenosis (update). *The spine journal : official journal of the North American Spine Society.* 2013;13(7):734-743.

29. Manchikanti L, Abdi S, Atluri S, et al. An update of comprehensive evidence-based guidelines for interventional techniques in chronic spinal pain. Part II: guidance and recommendations. *Pain Physician.* 2013;16(2 Suppl):S49-283.

30. Matz PG, Meagher RJ, Lamer T, et al. Guideline summary review: An evidence-based clinical guideline for the diagnosis and treatment of degenerative lumbar spondylolisthesis. *The spine journal : official journal of the North American Spine Society.* 2016;16(3):439-448.

31. Deer TR, Grider JS, Pope JE, et al. The MIST Guidelines: The Lumbar Spinal Stenosis Consensus Group Guidelines for Minimally Invasive Spine Treatment. *Pain practice : the official journal of World Institute of Pain.* 2019;19(3):250-274.

32. Rousing R, Jensen RK, Fruensgaard S, et al. Danish national clinical guidelines for surgical and nonsurgical treatment of patients with lumbar spinal stenosis. *European spine journal : official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society.* 2019;28(6):1386-1396.

33. Anderson DB, De Luca K, Jensen RK, et al. A critical appraisal of clinical practice guidelines for the treatment of lumbar spinal stenosis. *The Spine Journal.* 2020.

34. Chou R, Qaseem A, Owens DK, Shekelle P. Diagnostic imaging for low back pain: advice for high-value health care from the American College of Physicians. *Annals of internal medicine.* 2011;154(3):181-189.

35. UK NGC. Low back pain and sciatica in over 16s: assessment and management. 2016.

36. Downie A, Hancock M, Jenkins H, et al. How common is imaging for low back pain in primary and emergency care? Systematic review and meta-analysis of over 4 million imaging requests across 21 years. *British journal of sports medicine.* 2020;54(11):642-651.

37. Lee SY, Cho NH, Jung YO, Seo YI, Kim HA. Prevalence and Risk Factors for Lumbar Spondylosis and Its Association with Low Back Pain among Rural Korean Residents. *Journal of Korean Neurosurgical Society.* 2017;60(1):67-74.

38. Perera RS, Dissanayake PH, Senarath U, Wijayaratne LS, Karunanayake AL, Dissanayake VHW. Associations between disc space narrowing, anterior osteophytes and disability in chronic mechanical low back pain: a cross sectional study. *BMC Musculoskelet Disord.* 2017;18(1):193.

39. McDonagh MS, Selph SS, Buckley DI, et al. AHRQ Comparative Effectiveness Reviews. *Nonopioid Pharmacologic Treatments for Chronic Pain*. Rockville (MD): Agency for Healthcare Research and Quality (US); 2020.

40. Chou R, Hartung D, Turner J, et al. AHRQ Comparative Effectiveness Reviews. *Opioid Treatments for Chronic Pain*. Rockville (MD): Agency for Healthcare Research and Quality (US); 2020.

41. guideline NG193 N. Chronic pain (primary and secondary) in over 16s: assessment of all chronic pain and management of chronic primary pain. *Methods.* 2021.

42. Nicholas M, Vlaeyen JW, Rief W, et al. The IASP classification of chronic pain for ICD-11: chronic primary pain. *Pain.* 2019;160(1):28-37.

43. Holmberg T, Davidsen M, Thygesen LC, Krøll MJ, Tolstrup JS. Mortality among persons experiencing musculoskeletal pain: a prospective study among Danish men and women. *BMC musculoskeletal disorders.* 2020;21(1):1-12.

44. Roseen EJ, LaValley MP, Li S, et al. Association of back pain with all-cause and cause-specific mortality among older women: a cohort study. *Journal of general internal medicine.* 2019;34(1):90-97.

45. Smith D, Wilkie R, Croft P, Parmar S, McBeth J. Pain and mortality: mechanisms for a relationship. *Pain.* 2018;159(6):1112-1118.

CHAPTER TWO

Comparative Efficacy and Safety of Conservative Care for Pregnancy-Related Low Back Pain: A Systematic Review and Network Meta-analysis

Chapter Two has been published as:

Chen L., Ferreira ML, Beckenkamp PR, Caputo EL, Feng S, Ferreira PH. Comparative Efficacy and Safety of Conservative Care for Pregnancy-Related Low Back Pain: A Systematic Review and Network Meta-analysis. Phys Ther. 2021 Feb 4;101(2):pzaa200.

Statement from co-authors confirming authorship contribution of the PhD candidate

The co-authors of the paper "Comparative Efficacy and Safety of Conservative Care for Pregnancy-Related Low Back Pain: A Systematic Review and Network Meta-analysis" confirm that Lingxiao Chen has made the following contributions:

Conception and design of the research
Analysis and interpretation of the findings
Writing of the manuscript and critical appraisal of the content

In addition to the statements above, in cases where I am not the corresponding author of a published item, permission to include the published material has been granted by the corresponding author.

Lingxiao Chen Date: 21 July 2021

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.
Manuela Loureiro Ferreira Date: 21 July 2021

PTJ: Physical Therapy & Rehabilitation Journal | *Physical Therapy*, 2021;101:1–13
DOI: 10.1093/ptj/pzaa200
Advance access publication date November 19, 2020
Review

Comparative Efficacy and Safety of Conservative Care for Pregnancy-Related Low Back Pain: A Systematic Review and Network Meta-analysis

Lingxiao Chen, MMed[1,*], Manuela L Ferreira, PhD[1], Paula R. Beckenkamp, PhD[2], Eduardo L. Caputo, PhD[3], Shiqing Feng, PhD[4], Paulo H. Ferreira, PhD[2]

[1]Institute of Bone and Joint Research, Kolling Institute, Sydney Medical School, Faculty of Medicine and Health, University of Sydney, Sydney, Australia
[2]Discipline of Physiotherapy, Faculty of Health Sciences, University of Sydney, Sydney, Australia
[3]Postgraduate Program in Physical Education, Federal University of Pelotas, Pelotas, Brazil
[4]Department of Orthopaedics, Tianjin Medical University General Hospital, Tianjin, China
*Address all correspondence to Dr Chen at: lche4036@uni.sydney.edu.au

Abstract

Objective. More than one-half of pregnant women experience pregnancy-related low back pain (LBP). Pregnancy-related LBP greatly affects activities of daily life, and although many interventions have been proposed, the optimal treatment for pregnancy-related LBP remains unclear. The purpose of this study was to compare conservative care strategies on their efficacy and safety for women with pregnancy-related LBP through systematic review with pairwise meta-analysis and network meta-analysis.

Methods. MEDLINE, Embase, the Cochrane Library, AMED, CINAHL, PEDro, PsycINFO, and ClinicalTrials.gov were searched from inception to November 2019. Randomized controlled trials and observational controlled studies were included without restriction to language, sample size, or duration of follow-up. Two independent investigators extracted the data and assessed the risk of bias. The quality of evidence was evaluated through Grading of Recommendations Assessment, Development and Evaluation.

Results. Twenty-three studies were included in the qualitative synthesis (18 randomized controlled trials were included in the network meta-analysis). For women with LBP during pregnancy, progressive muscle relaxation therapy (mean difference = −3.96; 95% CI = −7.19 to −0.74; moderate-quality evidence) and Kinesio Taping (mean difference = −3.71; 95% CI = −6.55 to −0.87; low-quality evidence) reduced pain intensity (Visual Analog Scale, range = 0 to 10) compared with placebo. Moderate-quality evidence suggested that transcutaneous electrical nerve stimulation improved physical function (mean difference = −6.33; 95% CI = −10.61 to −2.05; Roland Morris Disability Questionnaire, range = 0–24) compared with placebo.

Conclusion. For patients with LBP during pregnancy, progressive muscle relaxation therapy and Kinesio Taping may help to decrease pain, and transcutaneous electrical nerve stimulation may improve physical function.

Impact. This review helps fill the gap in evidence regarding optimal treatment for pregnancy-related LBP.

Lay Summary. If you have LBP during pregnancy, your physical therapist has evidence to support the use of progressive muscle relaxation therapy and Kinesio Taping to help decrease pain and the use of transcutaneous electrical nerve stimulation to help improve physical function.

Keywords: Low Back Pain, Network Meta-Analysis, Pregnancy, Conservative Treatment

Received: January 14, 2020. **Accepted:** September 14, 2020

Introduction

Pregnancy-related low back pain (LBP) affects over one-half of women during pregnancy and is the leading cause for work absenteeism in this population.[1,2] Symptoms tend to persist 12 months postpartum for approximately 50% of women with pregnancy-related LBP, and up to 20% will present with residual symptoms even 3 years postpartum.[3,4] Two recent surveys conducted in the United Kingdom (including 176 pregnant women and 629 physiotherapists, respectively) indicated that conservative treatments, including exercise, manual therapy, yoga, and medications (eg, paracetamol and codeine/co-codamol), are the most commonly prescribed to treat pregnancy-related LBP.[5,6]

Despite many conservative strategies being commonly recommended for pregnancy-related LBP, little is known about the comparative effectiveness and safety of these approaches. A recent Cochrane Systematic Review has shown that land-based exercises can significantly decrease pain and disability compared with usual care for pregnancy-related LBP.[7] Likewise, 2 existing meta-analyses have demonstrated that manual therapy can significantly decrease pain and improve function compared with usual care or no treatment for pregnancy-related LBP and pelvic girdle pain. Compared with sham manipulation, however, no between-group differences were observed for manual therapy for the outcomes of pain and function.[8,9] Considering that previous reviews have only provided results for pairwise comparisons, the comparative effectiveness of these approaches (eg, exercise, manual therapy, medication) remains unknown. In addition, many of these studies[10] combined pregnancy-related LBP with pregnancy-related pelvic girdle pain, which might bias the results as these conditions present distinct prognosis.[11,12] Moreover, little is known about the safety of these interventions. A network meta-analysis can compare and rank multiple interventions simultaneously, filling the existing gap in the current literature.[13] We have conducted a network meta-analysis and systematic review of randomized controlled trials (RCTs) and observational studies to evaluate the effectiveness and safety of conservative care approaches for pregnancy-related LBP.

Methods

This review was prospectively registered on the International Prospective Register of Systematic Reviews (CRD42018093542),[14] and the report followed the Preferred Reporting Items for Systematic reviews and Meta-Analyses for Network Meta-Analyses.[15] The protocol is provided as supplementary material.

Data Sources and Searches

We searched MEDLINE (including MEDLINE Epub Ahead of Print, In-Process, & Other Non-Indexed Citations, MEDLINE Daily, and MEDLINE), Embase, the Cochrane Library, AMED, CINAHL, PEDro, PsycINFO, and Clinicaltrials.gov from their respective inception to April 2018. Search strategies for each database are provided as supplementary material. We updated our search in November 2019 to check for any newly published articles.

Study Selection

We included RCTs (parallel or crossover designs) and observational controlled studies without restriction to language, sample size, or duration of follow-up. Trials evaluating the effectiveness or efficacy of any conservative approach for pregnancy-related LBP were included. To maximally connect these interventions in the network meta-analysis, we included usual care and no treatment into the same node: control group. An intervention was defined as being usual care when either the authors described it as such (ie, Licciardone et al,[16] Shirazi et al,[17] and Gil et al[18]) or it encompassed minimal care (eg, Akmese and Oran[19] and Hensel et al[20]). Different sham interventions were grouped in 1 node: placebo. Two reviewers independently screened the titles and abstracts, which were imported into EndNote V8.2. Any disagreement was solved by discussion or arbitration by a third reviewer.

Data Extraction and Quality Assessment

Two reviewers independently extracted study characteristics (eg, sample size and location), patient characteristics (eg, age and body mass index), intervention characteristics, and outcome data from the included studies. Any disagreement was solved by discussion; otherwise, a third reviewer made the decision. Primary outcomes were pain intensity and physical function. Secondary outcomes were treatment withdrawal due to any reason and adverse events. We selected the data with the longest follow-up to include as many studies as possible in the network meta-analysis.

For the RCTs included in the network meta-analysis, a risk of bias tool including 13 items, based on the Cochrane Handbook for systematic reviews of interventions and recommended by the Cochrane Back and Neck Group, was used.[21] Studies were rated as presenting high risk of bias when 7 or more items were rated as having a high or unclear risk of bias.[21]

Data Synthesis and Analysis

Traditional pairwise meta-analyses through random-effect model with DerSimonian and Laird inverse-variance method for every direct comparison were performed.[22] Random-effect network meta-analyses with consistency model under the frequentist framework were performed to combine both direct and indirect comparisons.[13] We planned to perform meta-analyses for RCTs and observational controlled studies separately. Because only 1 observational study was identified, meta-analyses were only performed for RCTs. Continuous outcomes (pain intensity and physical function) were presented as mean differences (MD) or standardized mean differences (SMD), along with 95% confidence intervals (CIs) based on whether studies used the same score instruments.[22] The majority (75%, 9/12) of included studies used the Roland-Morris Disability Questionnaire to assess physical function in LBP during pregnancy. Thus, we selected MD rather than SMD, which meant we excluded 3 other studies in the meta-analyses because they reported physical function using different scales, to represent the outcome for an interpretable result. Dichotomous outcome (treatment withdrawal due to any reason) was presented as odds ratios with 95% CIs.[22] For pain intensity and physical function, the magnitude of effects was classified according to the recommendations from the American College of Physicians Clinical Practice Guideline (Suppl. Tab. 1) as: small/slight, moderate, or large/substantial.[23] All analyses were performed in Review manager V.5.3 (The Nordic Cochrane Centre, Copenhagen, Denmark, 2014) and Stata (StataCorp 2017. Stata Statistical

Software: Release 15.1. College Station, TX, USA LP). The statistical significance level was set at .05.

Sensitivity analyses were used for the primary outcomes by excluding studies with high risk of bias, except if the study was the only one to provide data for 1 intervention, studies from grey literature (ie, thesis), a study with a high rate of loss to follow-up (>70% vs < 20% in other included studies), and studies with suspected mixed populations. Three studies (during pregnancy: Gundermann S. 2013. Effectiveness of osteopathic treatment in pregnant women suffering from low back pain. A randomized controlled trial. [Unpublished D.O. thesis, Akademie für Osteopathie] and Röhrich K. 2014. Effectiveness of osteopathic treatment in pregnant women suffering from low back pain. A randomized controlled trial. [Unpublished D.O. thesis, Akademie für Osteopathie]; and postpartum: Schwerla et al[24]) including a mixed population of participants with LBP and pelvic girdle pain were pooled in the main analyses but excluded in subsequent sensitivity analyses. This was done to identify the impact of including a mixed clinical population on the magnitude of treatment effects. Meta-regression and subgroup analyses were planned but not performed as the number of studies in 1 comparison was <10 and insufficient information was provided, respectively.[25]

Network plots were drawn to describe the interventions for primary outcomes.[26] The Bucher method as a local method and design-by-treatment interaction model as a global method were used to assess the inconsistency.[13] If any inconsistency was found, the node-splitting method was performed to identify which comparisons contributed to inconsistency.[13] A comparison-adjusted funnel plot was used to assess publication bias if the number of included studies in the relevant outcome was >10.[26] The surface under the cumulative ranking curve and mean ranks with uncertainty intervals (2.5th percentile–97.5th percentile) were used to rank each intervention for each outcome.[26,27] We did not impute any missing data and only used available data to perform the meta-analyses. We used the Grading of Recommendations, Assessment, Development and Evaluations framework through Confidence in the Network Meta-analysis Internet application to evaluate the quality of evidence for the primary outcomes.[28,29]

Results

A total of 3837 articles were screened, and 23 studies were included in the qualitative synthesis (Fig. 1)[16–20,24,30–42] [4 unpublished studies: (1) Recknagel CRJ. 2007. Study on the effectiveness of osteopathic treatment for women with persistent post partum back pain. A randomized controlled trial. (Unpublished D.O. thesis, Akademie für Osteopathie); (2) Gundermann S. 2013. Effectiveness of osteopathic treatment in pregnant women suffering from low back pain. A randomized controlled trial. (Unpublished D.O. thesis, Akademie für Osteopathie); (3) Belz S. 2014. Effectiveness of osteopathic treatment in women with persistent non-specific low back pain after childbirth. A randomized controlled trial. (Unpublished master's thesis, Wiener Schule für Osteopathie); and (4) Röhrich K. 2014. Effectiveness of osteopathic treatment in pregnant women suffering from low back pain. A randomized controlled trial. (Unpublished D.O. thesis, Akademie für Osteopathie)]. Of the 23 included studies, 1 study was a non-RCT, 3 studies reported on interventions that could not be integrated into the network due to lack of connecting points, and 1 study provided insufficient outcome data and could not be combined with the remaining studies. Thus, 18 studies were included in the quantitative synthesis, 12 including women with LBP during pregnancy[16–20,32–33,38–40] (2 unpublished studies: Gundermann 2013 and Röhrich 2014, full citations above) and 6[34–37] including women with postpartum LBP (2 unpublished studies: Recknagel 2007 and Belz 2014, full citations above).

The characteristics of the included studies and baseline scores of pain intensity and physical function are presented in Table 1 and Suppl. Tab. 2, respectively. While most RCTs used a parallel-group design, 1 trial used a crossover design, and 1 study was a nonrandomized trial.[35,40] Overall, the baseline characteristics were in general similar across the included studies, although some of them (eg, details for the intervention and duration of follow-up) (Tab. 2) varied across studies. Thus, we performed a network meta-analysis considering transitivity assumption.[13] Among the 18 studies included in the network meta-analysis, 3[32,37,40] were rated as having high risk of bias (Suppl. Fig. 1).

Outcomes for LBP During Pregnancy

For pain intensity, assessed with the visual analogue scale (scored from 0 to 10, with lower scores indicating less pain), 12 RCTs[16–20,32–33,38–40] assessing 11 interventions and including 1276 participants were included in the network meta-analysis (Fig. 2) (2 unpublished studies: Gundermann 2013 and Röhrich 2014, full citations above). The network meta-analyses results showed that all interventions, except for exercise in combination with paracetamol and placebo interventions, significantly reduced pain intensity compared with the control group, with effect estimates ranging from −5.31 to −1.66 (Tab. 3). Progressive muscle relaxation therapy (MD = −3.96; 95% CI = −7.19 to −0.74; moderate-quality evidence due to imprecision and incoherence, indirect evidence) and Kinesio Taping (Kinesio Poland) (MD = −3.71; 95% CI = −6.55 to −0.87; low-quality evidence due to within-study bias and imprecision, 1 study) significantly reduced pain intensity compared with placebo. The results of the network meta-analysis suggest that progressive muscle relaxation therapy ranks first (Tab. 4). Overall, the quality of evidence was moderate or low due to within-study bias, imprecision, heterogeneity, and incoherence (Suppl. Tab. 3). The comparison-adjusted funnel plot (Suppl. Fig. 2) showed asymmetry, which might indicate publication bias.

For physical function, assessed with the Roland-Morris Disability Questionnaire (scored from 0 to 24, with lower scores indicating better function), 9 trials[16–18,20,33,38,40] assessing the effectiveness of 9 interventions in 1012 participants were included in the network meta-analysis (Fig. 2) (2 unpublished studies: Gundermann 2013 and Röhrich 2014, full citations above). All interventions resulted in improved physical function compared with the control group, with effect estimates (MD) ranging from −8.61 to −1.74. However, only Kinesio Taping in combination with paracetamol (MD = −7.53; 95% CI = −13.00 to −2.07; low-quality evidence due to imprecision and incoherence, indirect evidence), transcutaneous electrical nerve stimulation (MD = −8.61; 95% CI = −12.37 to −4.84; moderate-quality evidence due to imprecision and incoherence, 1 study), and rose oil (MD = −4.03; 95% CI = −7.49 to −0.58; moderate-quality

Table 1. Characteristics of Included Studies[a]

Time Related to Pregnancy	Study	No. of Participants	Study Design	Country	Duration of Follow-Up	Comparisons	Age, y	Parity	BMI	Gestational Age or Time Postpartum[b]
During	Abu et al,[39] (2017)[c]	86	RCT	Malaysia	6 wk	Exe + Par	29.4 (3.4)	**2 (4)**	23.6 (3.7)	17–28
						Par	29.4 (3.7)	**2 (4)**	23.8 (4.1)	
	Akmese and Oran[19] (2014)	73	RCT	Turkey	8 wk	PMR	20–35	N/A	19.5–30	12–24
						Con				
	Gil et al[18] (2011)	34	RCT	Brazil	8 wk	Exe	29 (5.2)	N/A	N/A	20–25
						Con	23.7 (3.9)			
	Gundermann[d] (2013)	41	RCT	Germany	6 wk	SMT	29 (3.2)	N/A	N/A	26.1 (3.2)
						Con	31 (4.1)			22.8 (4.4)
	Hensel et al[20] (2015)	400	RCT	United States	9 wk	SMT	23.99 (4.13)	N/A	25.51 (4.56)	≥30
						Pla	24.11 (4.1)		27.5 (6.44)	
						Con	24.7 (4.54)		27.54 (6.61)	
	Kalinowski and Krawulska[40] (2017)	106	Crossover	Poland	1 wk	Kin	29.5 (4.25)	2.42 (1.2)	22.3 (3.5)	Second or third trimester
						Pla				
	Kalus et al[26] (2008)	115	RCT	Australia	3 wk	BellyBra	N/A	N/A	N/A	28.2
						Tubigrip				29.2
	Kaplan et al[31] (2016)	66	RCT	Turkey	5 d	Kin + Par	24.3 (4.96)	1.06 (1.12)	26.5 (3.46)	10–30
						Par	25.09 (4.95)	1.41 (1.36)	25.6 (3.08)	
	Keskin et al[30] (2012)	79	RCT	Turkey	3 wk	TENS	29.1 (5)	**0 (1)**	N/A	**32 (1)**
						Par	29.7 (4.2)	**0 (1)**		**32 (1)**
						Exe	30.7 (4.3)	**0 (2)**		**32 (1)**
						Con	29.2 (4)	**1 (1)**		**32 (1)**
	Licciardone et al[16] (2010)	146	RCT	United States	9 wk	SMT	23.8 (5.5)	N/A	N/A	28–30
						Pla	23.7 (4.4)			
						Con	23.8 (5.2)			
	Mohseni Bandpei et al[32] (2010)	120	RCT	Iran	3 mo	Exe	24.71 (4.03)	N/A	23.79 (4.79)	17–22
						Con	25.11 (5.29)		23.15 (2.47)	
	Pekçetin et al[42] (2019)	124	RCT	Turkey	3 wk	Tel	28.53 (6.31)	N/A	N/A	<32
						Fac	28.18 (6.12)			
	Röhrich[e] (2014)	35	RCT	Germany	6 wk	SMT	32.7 (2.9)	N/A	N/A	24.1 (4.5)
						Con	30.3 (3.5)			20.5 (5.1)
	Sedaghati et al[30] (2007)	90	RCT	Iran	8 wk	Exe	23.3 (2.5)	N/A	24.1 (1.13)	20–22
						Con	23.3 (4.2)		24.3 (1.29)	
	Shirazi et al[17] (2017)	114	RCT	Iran	4 wk	Ros	27.7 (4.87)	0.6 (0.61)	26.5 (3.04)	22.1 (7.3)
						Pla	27.9 (4.32)	0.5 (0.62)	27.2 (5.55)	21.3 (6.78)
						Con	28.3 (3.75)	0.5 (0.62)	27.2 (5)	24.2 (5)
	Yan et al[35] (2014)	102	NRCT	China	12 wk	Exe	31.1	N/A	21.6 (3.47)	22–24
						Con	29.8		21.5 (3.36)	
After	Akbarzadeh et al[34] (2014)	100	RCT	Iran	2 wk	Cup	25 (4.2)	N/A	N/A	8 h
						Con	27 (3.8)			

(*Continued*)

Table 1. Continued

Time Related to Pregnancy	Study	No. of Participants	Study Design	Country	Duration of Follow-Up	Comparisons	Age, y	Parity	BMI	Gestational Age or Time Postpartum[b]
	Belz [f] (2014)	54	RCT	Germany	10 wk	SMT	33.8	1–3	N/A	N/A
						Con	34.3			
	Kamel et al[37] (2016)	45	RCT	Egypt	4 wk	SMT	37.4 (5.84)	N/A	27.02 (0.79)	N/A
						Pla	38.33 (4.38)		26.81 (0.52)	
						Con	37.46 (3.88)		26.82 (0.69)	
	Lee and Ko[36] (2015)	60	RCT	China	N/A	SMT	33.97 (2.93)	N/A	N/A	9–13 d
						Con	33.43 (3.54)			
	Mohamed et al[41] (2018)	30	RCT	Egypt	2 wk	Kin + Exe	24.15 (1.53)	N/A	28 (1.83)	3 mo
						Exe	23.42 (1.83)		27.84 (1.53)	
	Recknagel[g] (2007)	40	RCT	Germany	8 wk	SMT	34.5 (3.5)	N/A	N/A	3 mo
						Con	34.4 (5.0)			
	Schwerla et al[24] (2015)	80	RCT	Germany	8 wk	SMT	33.9 (4.4)	1–4	N/A	3–15 mo
						Con	33.3 (4.3)			

[a]Age, parity, body mass index (BMI), and gestational age or time postpartum are reported as mean (SD) or median (interquartile range if the data were available. Median (interquartile range) is indicated in bold. BMI = body mass index; Con = control group; Cup = cupping therapy; Exe = exercise; Fac = face-to-face ergonomic education; Kin = Kinesio Taping (Kinesio Poland); N/A = not applicable; NRCT = nonrandomized controlled trial; Par = paracetamol; Pla = placebo; PMR = progressive muscle relaxation; RCT = randomized controlled trial; Ros = rose oil; SMT = spinal manipulative treatment; Tel = telephone-supported ergonomic education; TENS = transcutaneous electrical nerve stimulation. [b]Reported in weeks unless otherwise indicated. This column refers to gestational age considered during pregnancy and time postpartum considered after pregnancy. [c]For age, parity, and BMI, all data were from the total population (59 participants with posterior pelvic pain and 86 participants with low back pain) because the article did not provide data only for low back pain. [d]Unpublished data: Gundermann S. 2013. Effectiveness of osteopathic treatment in pregnant women suffering from low back pain. A randomized controlled trial. Unpublished D.O. thesis, Akademie für Osteopathie. [e]Unpublished data: Röhrich K. 2014. Effectiveness of osteopathic treatment in pregnant women suffering from low back pain. A randomized controlled trial. Unpublished D.O. thesis, Akademie für Osteopathie. [f]Unpublished data: Belz S. 2014. Effectiveness of osteopathic treatment in women with persistent non-specific low back pain after childbirth. A randomized controlled trial. Unpublished master's thesis, Wiener Schule für Osteopathie. [g]Unpublished data: Recknagel CRJ. 2007. Study on the effectiveness of osteopathic treatment for women with persistent post partum low back pain. A randomized controlled trial. Unpublished D.O. thesis, Akademie für Osteopathie.

Table 2. Details of Interventions

Exercise and Spinal Manipulative Treatment in Studies Included in Network Meta-analyses	
Exercise During Pregnancy	
Gil et al[18] (2011)	1. Weekly 40-minute sessions for 8 weeks 2. Stretching of muscles of posterior chain—angle closure coxo-femoral and abduction of upper limbs and closing angle coxo-femoral with adduction of upper limbs
Keskin et al[33] (2012)	1. A home exercise program by a physical therapist as treatment modality. Program consisted of pelvic tilt exercises, stretching for lower extremity muscles, posture exercises, and mild isometric abdominal contractions. 2. Repeat each exercise 10×/session and to complete program twice daily for 3 wk.
Mohseni Bandpei et al[32] (2010)	1. Ergonomic advice and education using booklets and images 2. Exercise advice and education (1) 5 sessions of 20 min supervised program, in groups of 10, supervised by a physiotherapist and a midwife (2) using booklets and images 3. Home exercise program
Spinal manipulative treatment during pregnancy	
Gundermann[a] (2013)	Biweekly osteopathic treatment
Hensel et al[20] (2015)	Approx. 20 min treatment provided by physician board-eligible or certified by American Osteopathic Board of Neuro-musculoskeletal Medicine
Licciardone et al[16] (2010)	Study protocol included any of the following treatment modalities: soft tissue, myofascial release, muscle energy, and range-of-motion mobilization. Modalities were used in systematic manner within a protocol that enabled physician to identify and treat specific somatic dysfunctions in these anatomic regions: cervical, thoracic, and lumbar spine; thoracic outlet and clavicles; ribcage and diaphragm; and pelvis and sacrum.
Röhrich[b] (2014)	Biweekly osteopathic treatment
Postpartum spinal manipulative treatment	
Belz[c] (2014)	Biweekly osteopathic treatment
Kamel et al[37] (2016)	Central postero-anterior lumbar mobilization plus traditional treatment which consisted of Ultrasonic and Infra-red
Lee and Ko[36] (2015)	Each woman was asked to lie in a prone position. After body oil applied, massage therapist administered a 20-min reflexology session on each woman's back. Massage techniques involved effleurage, kneading, acupressure, and friction. Session conducted once every evening for 5 consecutive days.
Recknagel[d] (2007)	Biweekly osteopathic treatment
Schwerla et al[24] (2015)	Standard OMTh techniques (Glossary of Osteopathic Terminology) applied, including direct (high-velocity, low-amplitude; muscle energy; and myofascial release), indirect (functional techniques and balanced ligamentous tension), visceral, and cranial techniques
Details of control group and placebo for pain during pregnancy	
Control group	
Akmese and Oran[19] (2014)	Simply lie down and do nothing for 20 min 2×/d (morning and evening
Gil et al[18] (2011)	Usual prenatal care
Gundermann [a] (2013)	None
Hensel et al[20] (2015)	Usual prenatal care: completed study questionnaires but received no study interventions or additional time or interaction with treating physician
Keskin et al[33] (2012)	None
Licciardone et al[16] (2010)	Conventional prenatal care: conventional prenatal care during pregnancy exclusive of osteopathic manipulative treatment, which is generally considered complementary and alternative medicine therapy
Mohseni Bandpei et al[32] (2010)	None
Röhrich[b] (2014)	None
Shirazi et al[17] (2017)	Standard prenatal care
Placebo	
Hensel et al[20] (2015)	Placebo provided tactile and manual stimulation over same regions as intervention group. Ultrasound wand applied with circular, steady contact for approx. 2 min to each specified area, resulting in treatment duration similar to intervention group
Kalinowski and Krawulska[40] (2017)	Adhesive tape stuck without any tension on both sides of spine. Then, second tape used in transverse way at site where patient indicated pain
Licciardone et al[16] (2010)	Nonfunctional ultrasound therapy unit modified for research to provide visible and auditory cues that could potentially elicit placebo response plus conventional prenatal care
Shirazi et al[17] (2017)	Almond oil plus standard prenatal care

[a]Unpublished data: Gundermann S. 2013. Effectiveness of osteopathic treatment in pregnant women suffering from low back pain. A randomized controlled trial. Unpublished D.O. thesis, Akademie für Osteopathie. [b]Unpublished data: Röhrich K. 2014. Effectiveness of osteopathic treatment in pregnant women suffering from low back pain. A randomized controlled trial. Unpublished D.O. thesis, Akademie für Osteopathie. [c]Unpublished data: Belz S. 2014. Effectiveness of osteopathic treatment in women with persistent non-specific low back pain after childbirth. A randomized controlled trial. Unpublished master's thesis, Wiener Schule für Osteopathie. [d]Unpublished data: Recknagel CRJ. 2007. Study on the effectiveness of osteopathic treatment for women with persistent post partum low back pain. A randomized controlled trial. Unpublished D.O. thesis, Akademie für Osteopathie.

Table 3. Results of Network and Direct Evidence for Pain Intensity for Low Back Pain During Pregnancy[a]

PMR	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**−5.31 (−5.85to − 4.77)**
−0.25 (−4.55 to 4.04) Indirect evidence	Kin	N/A	N/A	N/A	N/A	N/A	N/A	N/A	**−3.71 (−3.09 to − 4.33)**	N/A
−0.53 (−5.41 to 4.36) Indirect evidence	−0.27 (−5.41 to 4.86) Indirect evidence	Kin + Par	N/A	N/A	N/A	**−1.92 (−2.82 to − 1.02)**	N/A	N/A	N/A	N/A
−1.45 (−5.33 to 2.43) Indirect evidence	−1.19 (−5.38 to 2.99) Indirect evidence	−0.92 (−5.01 to 3.17) Indirect evidence	TENS	N/A	N/A	**−1.00 (−1.74 to − 0.26)**	**−2.00 (−2.46 to − 1.54)**	N/A	N/A	**−3.00 (−3.45 to − 2.55)**
−1.45 (−6.29 to 3.40) Indirect evidence	−1.19 (−6.29 to 3.90) Indirect evidence	−0.92 (−4.99 to 3.15) Indirect evidence	−0.00 (−4.04 to 4.04) Indirect evidence	Exe + Par	N/A	**−1.00 (−1.63 to -0.37)**	N/A	N/A	N/A	N/A
−2.38 (−6.20 to 1.44) Indirect evidence	−2.12 (−5.96 to 1.71) Indirect evidence	−1.85 (−6.59 to 2.89) Indirect evidence	−0.93 (−4.63 to 2.77) Indirect evidence	−0.93 (−5.63 to 3.77) Indirect evidence	Ros	N/A	N/A	N/A	**−1.12 (−1.71 to − 0.53)**	**−3.4 (−3.98 to − 2.82)**
−2.45 (−6.37 to 1.47) Indirect evidence	−2.19 (−6.42 to 2.03) Indirect evidence	−1.92 (−4.83 to 0.99)	−1.00 (−3.87 to 1.87)	−1.00 (−3.84 to 1.84)	−0.07 (−3.81 to 3.67) Indirect evidence	Par	**−1.00 (−1.74 to − 0.26)**	N/A	N/A	**−2.00 (−2.74 to − 1.26)**
−2.58 (−6.13 to 0.97) Indirect evidence	−2.33 (−6.20 to 1.55) Indirect evidence	−2.05 (−6.04 to 1.94) Indirect evidence	−1.13 (−3.79 to 1.53)	−1.13 (−5.07 to 2.80) Indirect evidence	−0.20 (−3.55 to 3.14) Indirect evidence	−0.13 (−2.86 to 2.59)	Exe	N/A	N/A	**−3.15 (−6.13 to − 0.18)**
−3.65 (−6.81 to − 0.48) Indirect evidence	**−3.39 (−6.71 to − 0.07)** Indirect evidence	−3.12 (−7.34 to, 1.11) Indirect evidence	−2.20 (−5.20to 0.80) Indirect evidence	−2.20 (−6.37 to 1.98) Indirect evidence	−1.27 (−4.08 to 1.54) Indirect evidence	−1.20 (−4.26 to 1.86) Indirect evidence	−1.07 (−3.61 to 1.48) Indirect evidence	SMT	0.01 (−0.28 to 0.3)	**−1.46 (−2.52 to − 0.39)**
−3.96 (−7.19, to 0.74) Indirect evidence	**−3.71 (−6.55, to 0.87)**	−3.44 (−7.71 to 0.84) Indirect evidence	−2.52 (−5.59 to 0.56) Indirect evidence	−2.52 (−6.74 to 1.71) Indirect evidence	−1.59 (−4.16 to 0.99)	−1.52 (−4.65 to 1.61) Indirect evidence	−1.38 (−4.02 to 1.26) Indirect evidence	−0.32 (−2.04 to 1.41)	Pla	−0.96 (−2.08 to 0.16)
−5.31 (−8.13 to − 2.49)	**−5.06 (−8.29 to − 1.82)** Indirect evidence	**−4.78 (−8.77 to − 0.80)** Indirect evidence	**−3.86 (−6.52, −1.20)**	−3.86 (−7.80 to 0.07) Indirect evidence	**−2.93 (−5.51 to − 0.36)**	**−2.86 (−5.58 to − 0.14)**	**−2.73 (−4.88 to − 0.58)**	**−1.66 (−3.09 to − 0.24)**	−1.35 (−2.90 to 0.21)	Con

[a]The results from bottom-left are network evidence. The results from top-right are direct evidence. The statistically significant results are shown in bold type. The numbers are mean differences with 95% CIs. Con = control group; Exe = exercise; Kin = Kinesio Taping (Kinesio Poland); N/A = not applicable; Par = paracetamol; Pla = placebo; PMR = progressive muscle relaxation; Ros = rose oil; SMT = spinal manipulative treatment; TENS = transcutaneous electrical nerve stimulation.

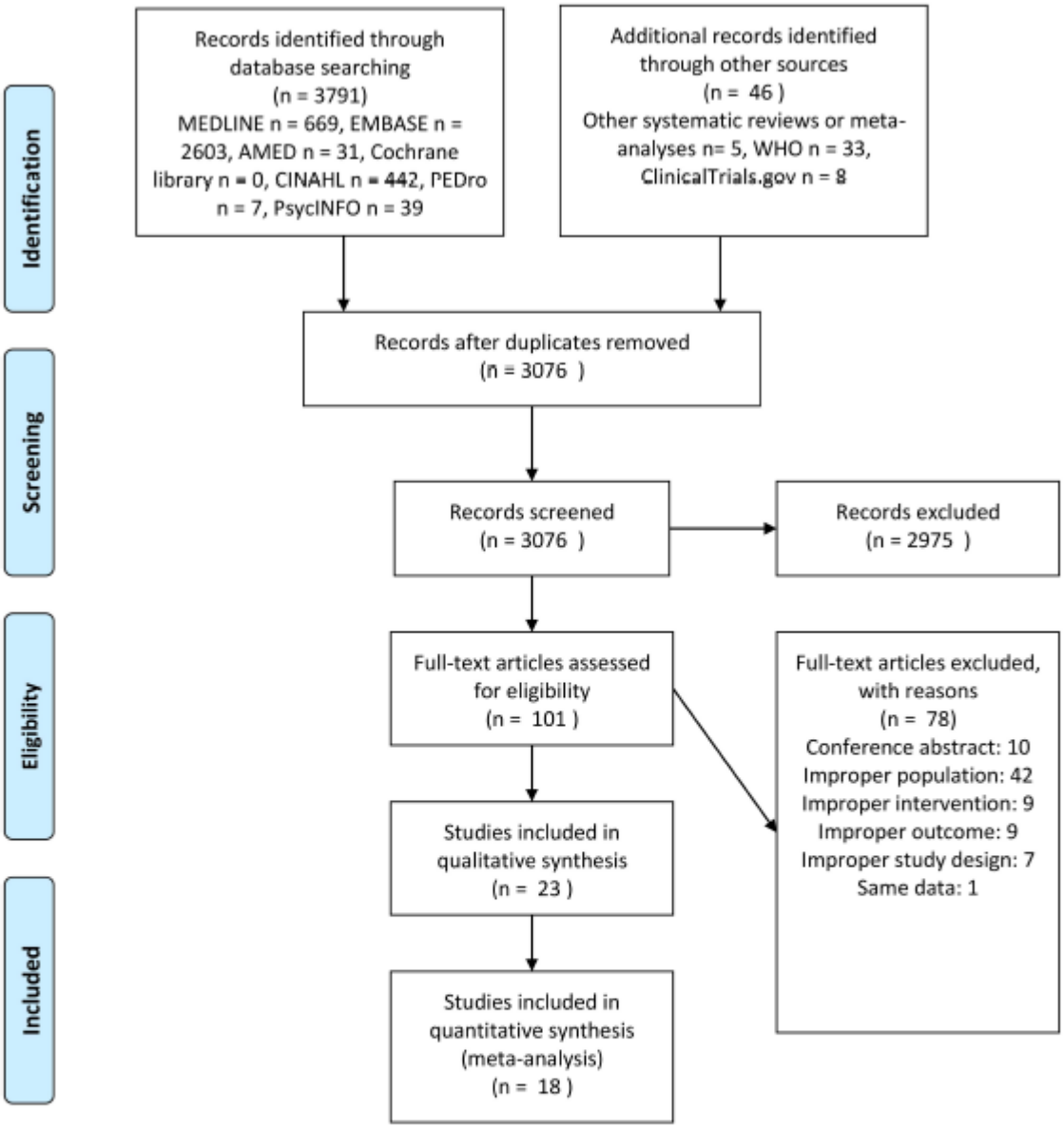


Figure 1. Evidence search and selection.

evidence due to imprecision and incoherence, 1 study) were statistically superior to the control groups (Suppl. Tab. 4). Transcutaneous electrical nerve stimulation significantly improved physical function compared with placebo (MD = −6.33; 95% CI = −10.61 to −2.05; moderate-quality evidence due to imprecision and incoherence, indirect evidence), spinal manipulative therapy (MD = −6.86; 95% CI = −11.07 to −2.66; moderate-quality evidence due to imprecision and incoherence, indirect evidence), paracetamol (MD = −5.00; 95% CI = −8.80 to −1.20; moderate-quality evidence due to imprecision, 1 study), and exercise (MD = −5.38; 95% CI = −9.16 to −1.61; moderate-quality evidence due to imprecision and incoherence, 1 study). The network meta-analysis ranked transcutaneous electrical nerve stimulation first (Suppl. Tab. 5). The quality of evidence was moderate or low due to within-study bias, imprecision, heterogeneity, and incoherence (Suppl. Tab. 3).

For treatment withdrawal due to any reason, 9 trials of 9 interventions with 1094 participants were included in the network meta-analysis. There were no statistical differences between any pairwise comparisons, and these interventions presented similar ranking (Suppl. Tab. 4 and 5). For adverse events, 7 studies reported at least 1 event, with the rate ranging from 1.6% to 15% in 1 intervention group (Suppl. Tab. 6).

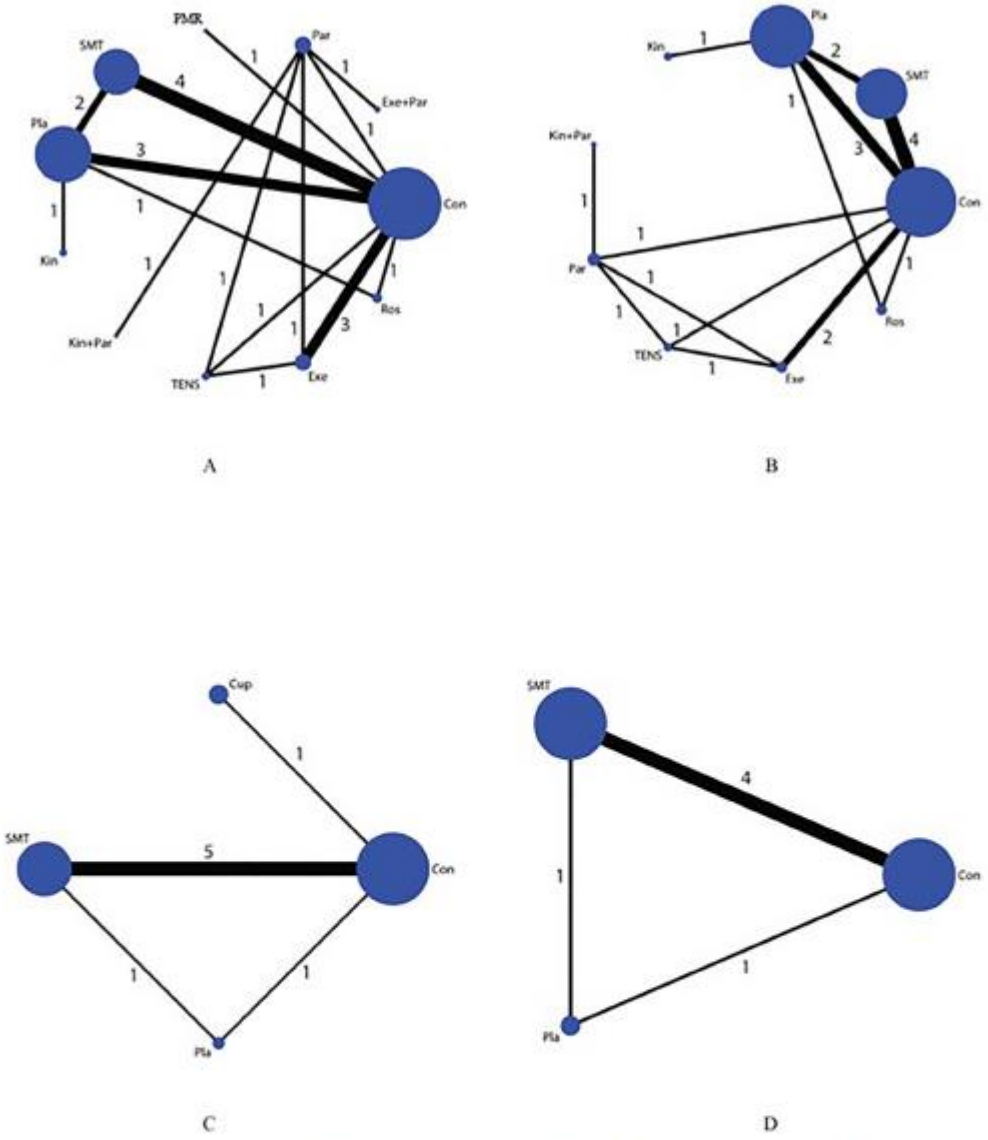


Figure 2. Network plots for all interventions. Each node represents 1 intervention, and the node size is proportional to the number of participants in each intervention. Each line represents a direct comparison between interventions, and the width of the line is proportional to the number of studies in each comparison. Con = control group; Cup = cupping therapy; Exe = exercise; Kin = Kinesio Taping (Kinesio Poland); Par = paracetamol; Pla = placebo; PMR = progressive muscle relaxation; Ros = rose oil; SMT = spinal manipulative treatment; TENS = transcutaneous electrical nerve stimulation. (A) Pain intensity for low back pain (LBP) during pregnancy. (B) Physical function for LBP during pregnancy. (C) Pain intensity for LBP postpartum. (D) Physical function for LBP postpartum.

Outcomes for LBP Postpartum

For pain intensity, measured using the visual analog scale (0–10), 6 trials[34–37] of 4 interventions and 378 participants were included in the network meta-analysis (Fig. 2) (2 unpublished studies: Recknagel 2007 and Belz 2014, full citations above). Spinal manipulative therapy (MD = −3.06; 95% CI = −4.09 to −2.03; moderate-quality due to imprecision, 5 studies) and cupping therapy (MD = −2.30; 95% CI = −4.51 to −0.09; moderate-quality evidence due to imprecision and heterogeneity, 1 study) significantly reduced pain intensity compared with control groups (Suppl. Tab. 4). Spinal manipulative therapy ranked first (Suppl. Tab. 5). The quality of evidence was moderate or very low due to within-study bias, imprecision, heterogeneity, and incoherence (Suppl. Tab. 3).

For physical function, assessed with the Oswestry Disability Index (0–50, with lower scores indicating better function) and the Pelvic Girdle Pain Questionnaire (0–60, with lower scores indicating better function), 4 trials[24,37] of 3 interventions and 218 participants were included in the network meta-analysis (Fig. 2) (2 unpublished studies: Recknagel 2007 and

Table 4. Ranking Results for the Outcome Pain Intensity for Low Back Pain During Pregnancy[a]

Rank	Name	SUCRA	Mean Rank (2.5th Percentile–97.5th Percentile)
First	Progressive muscle relaxation	83.2	2.7 (1–7)
Second	Kinesio Taping[b]	79.7	3.0 (1–8)
Third	Kinesio Taping + paracetamol	76.2	3.4 (1–9)
Fourth	Transcutaneous electrical nerve stimulation	64.1	4.6 (1–9)
Fifth	Exercise + paracetamol	62.3	4.8 (1–10)
Sixth	Rose oil	49.1	6.1 (2–10)
Seventh	Paracetamol	44.4	6.6 (3–10)
Eighth	Exercise	43.9	6.6 (3–10)
Ninth	Spinal manipulative treatment	25.6	8.4 (6–10)
Tenth	Placebo	20.1	9.0 (6–10)
Eleventh	Control	1.3	10.9 (10–11)

[a]SUCRA = surface under the cumulative ranking curve. [b]Kinesio Poland.

Belz 2014, full citations above). Spinal manipulative therapy (SMD = −2.20; 95% CI = −2.88 to −1.51; moderate-quality evidence due to heterogeneity, 4 studies) significantly improved physical function compared with the control group (Suppl. Tab. 4) and ranked first (Suppl. Tab. 5). The quality of evidence was moderate to very low due to within-study bias, imprecision, heterogeneity, and incoherence (Suppl. Tab. 5).

For treatment withdrawal due to any reason, 4 trials of 2 interventions were included in the pairwise meta-analysis. There was no significant difference between spinal manipulative therapy and control group (odds ratio = 1.02; 95% CI = 0.32 to 3.25). For adverse effect, only 1 study reported that some participants felt tired (Suppl. Tab. 6).

Outcomes From Studies Not Included in the Network Meta-analysis

For pain intensity, 4 studies were not included in the quantitative synthesis (Suppl. Tab. 7). These studies found that exercise and telephone-supported ergonomic education might be effective for women experiencing LBP during pregnancy compared with the control group and face-to-face ergonomic education respectively, and Kinesio Taping in combination with exercise might be effective for those reporting postpartum LBP compared with exercise. For physical function, 6 studies were not included in the quantitative synthesis (Suppl. Tab. 7). Overall, the results of these studies showed that progressive muscle relaxation therapy, exercise, BellyBra, and telephone-supported ergonomic education were effective for women experiencing LBP during pregnancy compared with control group, control group, Tubigrip group, and face-to-face ergonomic education, respectively, and Kinesio Taping in combination with exercise might be effective for those reporting postpartum LBP.

Sensitivity Analysis and Inconsistency Test

For pain intensity in patients with LBP during pregnancy, sensitivity analyses showed the result was stable (Suppl. Tab. 8). The inconsistency test identified inconsistency, and the node-splitting analysis was conducted showing statistical inconsistency for the placebo versus control group (P = .045), nearly statistical inconsistency for the paracetamol versus control group (P = .055), transcutaneous electrical nerve stimulation versus the control group (P = .054), exercise versus the paracetamol group (P = .053), and transcutaneous electrical nerve stimulation versus the exercise group (P = .053) (Suppl. Tab. 9 and 10). For the other 3 outcomes (physical function in patients with LBP during pregnancy, pain intensity in patients with LBP postpartum, and physical function in patients with LBP postpartum), sensitivity analyses did not alter the results (Suppl. Tab. 8) and no inconsistency was identified (Suppl. Tab. 9).

Discussion

The results of this systematic review and network meta-analyses found that for patients reporting LBP during pregnancy, progressive muscle relaxation therapy (moderate-quality evidence) and Kinesio Taping (low-quality evidence) statistically reduced pain intensity compared with placebo, whereas transcutaneous electrical nerve stimulation (moderate-quality evidence) statistically improved physical function compared with placebo. For patients reporting LBP in the postpartum period, spinal manipulative treatment (moderate-quality evidence) statistically reduced pain and improved physical function compared with the control group. Overall, treatment effects from all interventions were small to moderate except for some top-ranking interventions (eg, progressive muscle relaxation therapy, Kinesio Taping), which showed large (ie, >2 for the 0- to 10-point visual analogue scale and >5 for the 0- to 24-point Roland-Morris Disability Questionnaire) treatment effects. All interventions presented similar treatment withdrawal due to any reason and adverse events.

Previous pairwise meta-analyses showed similar results for the outcomes of pain and physical function in patients presenting with LBP during pregnancy for treatment comparisons such as manual therapy compared with control groups or sham or exercise, and exercise therapy in combination with usual care compared with usual care alone (Suppl. Tab. 11).[7–9] However, our results (MD = 3.65; 95% CI = 0.48 to 6.81) contradict a previous pairwise meta-analysis (SMD = −0.77; 95% CI = −1.22 to −0.32) that showed that manual therapy significantly reduced pain compared with relaxation techniques (Suppl. Tab. 11).[9] We believe the main reason for this discrepancy was due to inclusion of different studies in these 2 reviews. The previous systematic review included 2 studies that recruited patients diagnosed with depression in addition to LBP, whereas ours included studies that recruited patients diagnosed with LBP only (although many included studies did not report whether they excluded patients with severe mental diseases). Several studies have shown that depression is a predictor of pain

and associated with poor treatment outcome in adults with chronic LBP.[43,44] Thus, the presence of depression in pregnant women may affect response to treatment, leading to the inconsistency in results between the 2 reviews. But there were insufficient data in the original trials to explore that relationship in the current review. Only 1 previous pairwise meta-analysis assessed safety outcomes for manual therapy compared with sham or exercise. This review found similar results to ours, with no significant differences between manual therapy and sham or exercise (Suppl. Tab. 11).[9] For patients with postpartum LBP, 1 previous pairwise meta-analysis indicated that osteopathic manipulative therapy significantly reduced pain and improved physical function compared with the control group, which is in line with our results (Suppl. Tab. 11).[8]

Currently, there are no clinical practice guidelines for pregnancy-related LBP that exclude pelvic girdle pain. Current clinical guidelines for the treatment of non–pregnancy-related LBP or pelvic girdle pain including LBP endorse interventions that were highly ranked in our review, such as spinal manipulative therapy.[45–48] However, little evidence supported Kinesio Taping and transcutaneous electrical nerve stimulation,[48] with the National Institute for Health and Clinical Excellence guideline not recommending transcutaneous electrical nerve stimulation for patients with LBP or sciatica.[45] Possible reasons for the disagreements between previous guidelines and our study include different types of patients (these guidelines excluded pregnant women). For Kinesio Taping, the previous 2 studies were conducted in different settings (Brazil and Spain in LBP reviews vs Poland and Turkey in the current review).[49,50] For transcutaneous electrical nerve stimulation, several studies demonstrated its effects in reducing pain in pregnant women.[51–53] Thus, Kinesio Taping–related therapies and transcutaneous electrical nerve stimulation might be effective in pregnancy-related LBP; however, further high-quality trials should be performed assessing the effectiveness of these 2 therapies. Finally, current guidelines are insufficient regarding recommendations for pregnancy-related LBP that exclude pelvic girdle pain, and therefore the results from our review could be a source of information to guide management for this clinical group.

This network meta-analysis presents some limitations. First, 3 of the included studies presented high risk of bias. However, 2 of them[37,40] were the only studies to provide data for 1 intervention (Kinesio Taping for women during pregnancy and placebo for women postpartum), and we could only exclude 1 study[32] in the outcome of pain intensity for LBP during pregnancy. Therefore, the results of these interventions should be interpreted cautiously. Second, some baseline characteristics (eg, pain intensity and physical function); the duration of follow-up; and the details for the intervention exercise, spinal manipulative therapy, placebo, and control group (eg, usual care definition, usual care, and no treatment were grouped together) varied across studies. We did not perform meta-regression or subgroup analyses to explore their possible different effects because of the limited number of included studies. Third, there were some inconsistencies (35.7% of all comparison including direct and indirect evidences) in the outcome of pain intensity for LBP during pregnancy, which might influence the robustness of our result. Fourth, some interventions (eg, Kinesio Taping in combination with exercise and education) could not be integrated into the network meta-analyses because we lack the essential studies to connect these interventions. Finally, for postpartum LBP, the included studies only reported data on the effectiveness of 4 interventions, which is significantly less than the number of interventions used in clinical practice.

Progressive muscle relaxation therapy and Kinesio Taping may be helpful to reduce pain, while transcutaneous electrical nerve stimulation may improve physical function in patients with LBP during pregnancy. There is urgent need for more high-quality RCTs in the field, especially for interventions that have been investigated in only 1 or 2 studies (eg, progressive muscle relaxation therapy, Kinesio Taping, and transcutaneous electrical nerve stimulation) and patients with postpartum LBP considering the number of interventions and studies are small.

Author Contributions

Concept/idea/research design: L. Chen, M. Ferreira, P. Beckenkamp, P. Ferreira

Writing: L. Chen, P. Beckenkamp, E. Caputo, S. Feng

Data collection: L. Chen, E. Caputo

Data analysis: L. Chen

Project management: L. Chen, M. Ferreira, P. Ferreira

Consultation (including review of manuscript before submitting): P. Ferreira

Acknowledgments

The authors thank Ms Sai Zhao for providing methodological help, Dr Karin Röhrich and Dr Florian Schwerla for providing grey literature, Dr Yulia Ulyannikova for providing suggestions about how to develop search strategies, and Dr Claudia Rabenberg and Dr Fereshteh Pourkazemi for translating articles (German and Persian, respectively) and extracting data.

Funding

There are no funders to report for this study. M. Ferreira holds a National Health and Medical Research Council of Australia (NHMRC) Fellowship (2018–2021) and the University of Sydney SOAR Fellowship (2019–2020). E. Caputo is financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior.—Brasil (CAPES)—Finance Code 001.

Systematic Review Registration

This review was prospectively registered in the international prospective register of systematic reviews (CRD42018093542). The report followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Network Meta-Analyses.

Disclosures

All authors completed the ICMJE Form for Disclosure of Potential Conflicts of Interest and reported no conflicts of interest.

References

1. Ostgaard HC, Andersson GB. Previous back pain and risk of developing back pain in a future pregnancy. *Spine*. 1991;16:432–436.
2. Sencan S, Ozcan-Eksi EE, Cuce I, Guzel S, Erdem B. Pregnancy-related low back pain in women in Turkey: prevalence and risk factors. *Ann Phys Rehabil Med*. 2018;61:33–37.

3. Noren L, Ostgaard S, Johansson G, Ostgaard HC. Lumbar back and posterior pelvic pain during pregnancy: a 3-year follow-up. *Eur Spine J*. 2002;11:267–271.
4. Padua L, Caliandro P, Aprile I, et al. Back pain in pregnancy: 1-year follow-up of untreated cases. *Eur Spine J*. 2005;14:151–154.
5. Bishop A, Holden MA, Ogollah RO, Foster NE. Current management of pregnancy-related low back pain: a national cross-sectional survey of UK physiotherapists. *Physiotherapy*. 2016;102:78–85.
6. Marlene Sinclair CC, McCullough JEM, Hughes C, Dianne Liddle S. How do women manage pregnancy-related low back and/or pelvic pain? Descriptive findings from an online survey. *Evid Based Midwifery*. 2014;12:76–82.
7. Liddle SD, Pennick V. Interventions for preventing and treating low-back and pelvic pain during pregnancy. *Cochrane Database Syst Rev*. 2015;2015:CD001139.
8. Franke H, Franke JD, Belz S, Fryer G. Osteopathic manipulative treatment for low back and pelvic girdle pain during and after pregnancy: a systematic review and meta-analysis. *J Bodyw Mov Ther*. 2017;21:752–762.
9. Hall H, Cramer H, Sundberg T, et al. The effectiveness of complementary manual therapies for pregnancy-related back and pelvic pain: a systematic review with meta-analysis. *Medicine*. 2016;95:e4723.
10. Davenport MH, Marchand AA, Mottola MF, et al. Exercise for the prevention and treatment of low back, pelvic girdle and lumbopelvic pain during pregnancy: a systematic review and meta-analysis. *Br J Sports Med*. 2019;53:90–98.
11. Van De Pol G, Van Brummen HJ, Bruinse HW, Heintz AP, Van Der Vaart CH. Pregnancy-related pelvic girdle pain in the Netherlands. *Acta Obstet Gynecol Scand*. 2007;86:416–422.
12. Vermani E, Mittal R, Weeks A. Pelvic girdle pain and low back pain in pregnancy: a review. *Pain Pract*. 2010;10:60–71.
13. White IR. Network meta-analysis. *Stata J*. 2015;15:951–985.
14. Chen L, Caputo EL, Beckenkamp P, Ferreira M, Ferreira P. The comparative efficacy and safety of conservative care for pregnancy-related low back pain: a protocol for a network meta-analysis and systematic review. *PROSPERO*. 2018;CRD42018093542. Accessed October 31, 2020 http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42018093542.
15. Hutton B, Salanti G, Caldwell DM, et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. *Ann Intern Med*. 2015;162:777–784.
16. Licciardone JC, Buchanan S, Hensel KL, King HH, Fulda KG, Stoll ST. Osteopathic manipulative treatment of back pain and related symptoms during pregnancy: a randomized controlled trial. *Am J Obstet Gynecol*. 2010;202:43.e1–43.e438.
17. Shirazi M, Mohebitabar S, Bioos S, et al. The effect of topical *Rosa damascena* (rose) oil on pregnancy-related low back pain: a randomized controlled clinical trial. *J Evid Based Complementary Altern Med*. 2017;22:120–126.
18. Gil VFB, Osis MJD, Faúndes A. Lombalgia durante a gestação: eficácia do tratamento com Reeducação postural global (RPG). *Fisioterapia e Pesquisa*. 2011;18:164–170.
19. Akmese ZB, Oran NT. Effects of progressive muscle relaxation exercises accompanied by music on low back pain and quality of life during pregnancy. *J Midwifery Womens Health*. 2014;59:503–509.
20. Hensel KL, Buchanan S, Brown SK, Rodriguez M, Cruser dA. Pregnancy research on osteopathic manipulation optimizing treatment effects: The PROMOTE study. *Am J Obstet Gynecol*. 2015;212:108.e1–108.e1089.
21. Furlan AD, Malmivaara A, Chou R, et al. 2015 updated method guideline for systematic reviews in the Cochrane back and neck group. *Spine*. 2015;40:1660–1673.
22. Higgins JPT GS. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]*. The Cochrane Collaboration, 2011. Accessed October 31, 2020 https://handbook-5-1.cochrane.org/.
23. Chou R, Deyo R, Friedly J, et al. Systemic pharmacologic therapies for low back pain: a systematic review for an American College of Physicians clinical practice guideline. *Ann Intern Med*. 2017;166:480–492.
24. Schwerla F, Rother K, Rother D, Ruetz M, Resch KL. Osteopathic manipulative therapy in women with postpartum low back pain and disability: a pragmatic randomized controlled trial. *J Am Osteopath Assoc*. 2015;115:416–425.
25. White IR. Multivariate random-effects meta-regression: updates to mvmeta. *Stata J*. 2011;11:255–270.
26. Chaimani A, Higgins JP, Mavridis D, Spyridonos P, Salanti G. Graphical tools for network meta-analysis in STATA. *PloS One*. 2013;8:e76654.
27. Veroniki AA, Straus SE, Rucker G, Tricco AC. Is providing uncertainty intervals in treatment ranking helpful in a network meta-analysis? *J Clin Epidemiol*. 2018;100:122–129.
28. Salanti G, Del Giovane C, Chaimani A, Caldwell DM, Higgins JPT. Evaluating the quality of evidence from a network meta-analysis. *PloS One*. 2014;9:e99682.
29. CINeMA: *Confidence in Network Meta-Analysis [Software]. Institute of Social and Preventive Medicine*, University of Bern, 2017. Accessed November 28, 2018. Available from cinema.ispm.ch.
30. Sedaghati P, Ziaee V, Ardjmand A. The effect of an ergometric training program on pregnants' weight gain and low back pain. *Gazzetta Medica Italiana*. 2007;166:209–213.
31. Kalus SM, Kornman LH, Quinlivan JA. Managing back pain in pregnancy using a support garment: a randomised trial. *BJOG*. 2008;115:68–75.
32. Mohseni Bandpei MA, Åhmadshirvani M, Fakhri M, Rahmani N. The effect of an exercise program and ergonomic advices on treatment of pregnancy-related low back pain: a randomized controlled clinical trial. *Mazandaran Univ Med Sci*. 2010;20:10–19.
33. Keskin EA, Onur O, Keskin HL, Gumus II, Kafali H, Turhan N. Transcutaneous electrical nerve stimulation improves low back pain during pregnancy. *Gynecol Obstet Invest*. 2012;74:76–83.
34. Akbarzadeh M, Ghaemmaghami M, Yazdanpanahi Z, Zare N, Azizi A, Mohagheghzadeh A. The effect dry cupping therapy at acupoint BL23 on the intensity of postpartum low back pain in primiparous women based on two types of questionnaires, 2012: a randomized clinical trial. *Int J Community Based Nurs Midwifery*. 2014;2:112–120.
35. Yan C-F, Hung Y-C, Gau M-L, Lin K-C. Effects of a stability ball exercise programme on low back pain and daily life interference during pregnancy. *Midwifery*. 2014;30:412–419.
36. Lee HJ, Ko YL. Back massage intervention for relieving lower back pain in puerperal women: a randomized control trial study. *Int J Nurs Pract*. 2015;21(suppl 2):32–37.
37. Kamel DM, Raoof NA, Tantawy SA. Efficacy of lumbar mobilization on postpartum low back pain in Egyptian females: a randomized control trial. *J Back Musculoskelet Rehabil*. 2016;29:55–63.
38. Kaplan S, Alpayci M, Karaman E, et al. Short-term effects of Kinesio taping in women with pregnancy-related low back pain: a randomized controlled clinical trial. *Med Sci Monit*. 2016;22:1297–1301.
39. Abu MA, Abdul Ghani NA, Shan LP, et al. Do exercises improve back pain in pregnancy? *Horm Mol Biol Clin Investig*. 2017;32. doi: 10.1515/hmbci-2017-0012.
40. Kalinowski P, Krawulska A. Kinesio taping vs. placebo in reducing pregnancy-related low back pain: a cross-over study. *Med Sci Monit*. 2017;23:6114–6120.
41. Mohamed EA, El-Shamy FF, Hamed H. Efficacy of kinesiotape on functional disability of women with postnatal back pain: a randomized controlled trial. *J Back Musculoskelet Rehabil*. 2018;31:205–210.
42. Pekçetin S, Özdinç S, Ata H, Can HB, Elter K. Effect of telephone-supported ergonomic education on pregnancy-related low back pain. *Women Health*. 2019;59:294–304.
43. Adnan R, Van Oosterwijck J, Cagnie B, et al. Determining predictive outcome factors for a multimodal treatment program in low

back pain patients: a retrospective cohort study. *J Manipulative Physiol Ther.* 2017;40:659–667.
44. Jung SH, Kwon OY, Yi CH, et al. Predictors of dysfunction and health-related quality of life in the flexion pattern subgroup of patients with chronic lower back pain: The STROBE study. *Medicine.* 2018;97:e11363.
45. National Institute for Health and Clinical Excellence. *Low back pain and sciatica in over 16s: assessment and management.* NICE guideline (NG59); 2016. https://www.nice.org.uk/guidance/ng59.
46. Clinical practice guideline: management of pelvic girdle pain in pregnancy and post-partum. Chartered physiotherapists Women's health and continence and Directorate of Strategy and Clinical Programmes Health Service Executive. 2014. https://www.hse.ie/eng/services/publications/clinical-strategy-and-programmes/management-of-pelvic-girdle-pain-in-pregnancy-and-post-partum.pdf.
47. Clinton SC, Newell A, Downey PA, Ferreira K. Pelvic girdle pain in the antepartum population: physical therapy clinical practice guidelines linked to the international classification of functioning, disability and health from the Section on Women's Health and the Orthopaedic Section of the American Physical Therapy Association. *J Womens Health Phys Therap.* 2017;41:102–125.
48. Qaseem A, Wilt TJ, McLean RM, Forciea M, for the Clinical Guidelines Committee of the American College of Physicians. Noninvasive treatments for acute, subacute, and chronic low back pain: a clinical practice guideline from the American College of Physicians. *Ann Intern Med.* 2017;166:514–530.
49. Castro-Sanchez AM, Lara-Palomo IC, Mataran-Penarrocha GA, Fernandez-Sanchez M, Sanchez-Labraca N, Arroyo-Morales M. Kinesio taping reduces disability and pain slightly in chronic non-specific low back pain: a randomised trial. *J Physiother.* 2012;58:89–95.
50. Parreira Pdo C, Costa Lda C, Takahashi R, et al. Kinesio taping to generate skin convolutions is not better than sham taping for people with chronic non-specific low back pain: a randomised trial. *J Physiother.* 2014;60:90–96.
51. Feng X, Ye T, Wang Z, et al. Transcutaneous acupoint electrical stimulation pain management after surgical abortion: a cohort study. *Int J Surg.* 2016;30:104–108.
52. Kayman-Kose S, Arioz DT, Toktas H, et al. Transcutaneous electrical nerve stimulation (TENS) for pain control after vaginal delivery and cesarean section. *J Matern Fetal Neonatal Med.* 2014;27:1572–1575.
53. Shahoei R, Shahghebi S, Rezaei M, Naqshbandi S. The effect of transcutaneous electrical nerve stimulation on the severity of labor pain among nulliparous women: a clinical trial. *Complement Ther Clin Pract.* 2017;28:176–180.

Supplementary Material

Items
Protocol registration on PROSPERO
Search strategy
Supplementary Table 1. Magnitude of Effects
Supplementary Table 2. Baselines of primary outcomes in included studies
Supplementary Figure 1. Risk of bias of included randomized controlled trials in network meta-analyses
Supplementary Table 3. GRADE results
Supplementary Figure 2. Comparison-adjusted funnel plot for pain during pregnancy
Supplementary Table 4. Results from direct and network evidence
Supplementary Table 5. Rank results
Supplementary Table 6. Adverse events of included studies
Supplementary Table 7. Effect estimates from studies not included in the network meta-analysis
Supplementary Table 8. Sensitivity analyses
Supplementary Table 9. Loop and global inconsistency
Supplementary Table 10. Results of node-splitting method for pain during pregnancy
Supplementary Table 11. Previous meta-analyses

Protocol registration on PROSPERO

Deviations in the formal study

Some of the planned secondary outcomes (health-related quality of life, global impression of recovery, and work absenteeism) were not reported because the data from the included studies could not be incorporated into meta-analysis. We added more sensitivity analyses to test the robustness of the results.

Comparative efficacy and safety of conservative care for pregnancy-related low back pain: protocol for a network meta-analysis and systematic review

Introduction

Pregnancy-related low back pain (LBP) is common in pregnant women, and the prevalence of it is about 50% (range 25% to 90%) (1-3). It is also the most common complaint about musculoskeletal disorders in pregnant women (4). Some women (about 33%) suffer severe pain, which reduces their quality of life (2). Worse, the symptoms of pregnancy-related LBP will still exist in 50% of pregnant women one year postpartum and 20% three years after delivery (5, 6). In addition, it ranks first in all reasons for working pregnant women to ask for sick leave (7).

Conservative treatments should be preferred considering the potential risk of invasive and surgical interventions (8). A 2014 UK survey indicated that paracetamol, codeine/co-codamol, NSAIDs, osteopathy, reflexology and chiropractic treatment were the popular options (9). A lot of other options could also be chosen, such as massage, acupuncture, Yoga, exercise, and cyclobenzaprine (8, 10). Previous meta-analyses and

systematic reviews were pairwise and therefore could notprovide comparative efficacy and safety of these conservative interventions (10-15). A network meta-analysis could overcome the limitation, by simultaneously comparing all these interventions and making a hierarchy of all interventions in each outcome (16). We aim to perform a network meta-analysis and systematic review to compare efficacy and safety of conservative care for pregnancy-related LBP.

Methods and Analysis

Criteria for Considering Studies for this Review

The protocol is written based on the Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P)(17). Any revise on this protocol will be updated in the PROSPERO registration.

Types of Participants

Studies involving pregnant women who are diagnosed with pregnancy-related low LBP will be included. We will exclude studies involving patients with malignancy, trauma, vertebral fracture, infection, and inflammatory disorders. For studies including pregnancy-related pelvic girdle pain or other diseases, we will include the study if we can extract the data for pregnancy-related LBP or at least 80% of pregnant women are diagnosed with pregnancy-related LBP.

Types of Interventions

Any pharmacological or non-pharmacological, non-invasive options for pregnancy-related LBP will be included. The examples of pharmacological options are acetaminophen, NSAIDs, opioids, tramadol and tapentadol, antidepressants, skeletal

muscle relaxants, benzodiazepines, corticosteroids, and antiseizure medications. For non-pharmacological, non-invasive options, exercise, Taichi, Yoga, mindfulness-based stress reduction, psychological therapy, multidisciplinary rehabilitation, acupuncture, spinal manipulation and transcutaneous electrical nerve stimulation will be included. The comparison group could be no treatment, usual care, another active option or combination of options.

Outcome Measures

The outcome will be divided into two parts: during the pregnancy and postpartum. If included studies report one outcome in several time points, we will choose the data from the longest duration of follow-up to perform the primary analysis. For postpartum outcome, we will perform a subgroup-analysis.

Primary Outcomes

1. Physical function, commonly measured by Oswestry disability index (ODI), Roland Morris disability questionnaire (RMDQ) and core outcome measures index (COMI) (18). Other rating scales will be included if they have been proposed in peer-reviewed journals.

2. Pain intensity, commonly measured by numeric rating scale (NRS) and the visual analog scale (VAS) (19, 20). Other rating scales will also be included if they have been proposed in peer-reviewed journals.

Secondary Outcomes

Health-related quality of life (HRQOL), commonly measured by SF-36, EQ-5D, Nottingham health profile (NHP) and SF-12(18). SF-36, NHP, and SF-12 could be

mapped into EQ-5D (21). Other rating scales will be included if they have been proposed in peer-reviewed journals.

Global impression of recovery, measured by the percentage of the patients who satisfy the recovery.

Work absenteeism, measured by the length of sick leave.

Adverse effect, measured by the percent of any adverse event.

Treatment withdrawal due to any reason, measured by percent of the patients who drop out.

Types of Studies

For the efficacy outcomes (physical function, pain intensity, HRQOL, global impression of recovery and work absenteeism), only randomized controlled trials (RCT) will be included. For safety outcomes (adverse effect and treatment withdrawal due to any reason), RCTs and non-randomized studies (NRCT) with a control group will be included. For cross-over studies, only data before wash-out period will be used. For cluster randomized trials, we will extract data which is adjusted for clustering. If these data are unavailable, we will extract original data and adjust them (22, 23).

Search Strategy

Electronic searches

The following databases will be searched: MEDLINE, Embase, the Cochrane Library, AMED, CINAHL, PEDro, PsycINFO and Clinicaltrials.gov.

Reference Lists and Other Sources

Reference lists of all included studies, relevant systematic reviews and meta-analyses, and guidelines will be screened to check whether there is a study to be included.

Identification and Selection of Studies

Two reviewers will independently screen the titles and abstracts of the articles from the search. Before the formal screening, we will perform a pilot test: we will randomly choose 50 citations using a random number table to confirm we have enough inter-rater agreement (at least 80%). Discussion will solve any disagreement. Otherwise, a third reviewer will make a decision. If some articles' necessary information is missing, we will contact the corresponding author or first author.

Data Extraction

Two reviewers will independently extract data from included studies using a standardized data extraction form. Similarly, a pilot test will be performed before the formal extraction. We will randomly choose five articles using a random number table to confirm we have enough inter-rater agreement (at least 80%). Any disagreement will be solved by discussion. Otherwise, a third reviewer will make a decision. The following data will be extracted from each included study based on the recommendations from previous studies (18, 24).

Study characteristics, such as year of study publication, first author, journal, sample size, study funding, and location.
Patient characteristics, such as age, gender, including and excluding criteria, diagnostic criteria, comorbidities and previous treatment.
Intervention characteristics.
Primary and secondary outcomes.

Measurement of Treatment Effect

Relative treatment effects

Continues outcomes: If the studies use the same rating scale, we will use mean difference (MD) with its 95% confidence interval (CI). If different rating scales, standardized mean difference (SMD) with its 95% CI will be used.
Dichotomous outcomes: odds ratio (OR) with its 95% CI will be used.

Relative treatment ranking

The surface under the cumulative ranking curve (SUCRA) and mean ranks with uncertain interval will be used to rank each intervention for each outcome (25).

Dealing with missing outcome data and missing statistics

For continuous outcomes, if the study only reports standard error (SE), P value or CI, we will convert them into standard deviance (SD) (23). If the study reports median and interquartile range (IQR), we will calculate SD through divide IQR by 1.35 and consider median equals mean (23). If relevant information is in the figure, we will extract the data from the figure. If no data could be obtained, we will contact corresponding or first author. If fail, we will perform available data analysis only (23).
For dichotomous outcomes, firstly, we will try to contact corresponding or first author. If fail, similarly, we will perform available data analysis only (23).

Risk of bias assessment

Two reviewers will independently assess the risk of bias in included studies. Any disagreement will be solved by discussion. Otherwise, a third reviewer will make a

decision. We will contact corresponding or first author to obtain further information if the third reviewer thinks it is necessary.

For RCT, risk of bias tool will be used based on Cochrane Handbook for systematic reviews for interventions and the recommendation from Cochrane Back and Neck Group (23, 24). The tool has 13 items, which is: 1. Random sequence generation; 2. Allocation concealment; 3. Blinding of participants; 4. Blinding of personnel/ care providers; 5. Blinding of outcome assessor; 6. Incomplete outcome data; 7. Selective Reporting; 8. Group similarity at baseline; 9. Co-interventions 10. Compliance; 11. Intention-to-treat-analysis; 12. Timing of outcome assessments; 13. Other Bias. For the item 13, we will mainly focus on whether the study received commercial funding. For each item, we will rate it as low risk of bias, unclear risk of bias or high risk of bias. If 7 or more items are rated as low risk of bias and the study has no serious flaws, we will rate the study as low risk of bias, or we will rate the study as high risk of bias (26, 27).

For non-randomized trials, Risk Of Bias In Non-randomised Studies - of Interventions (ROBINS-I) tool will be used (28). The tool has seven domains: 1. Bias due to confounding; 2. Bias in selection of participants into the study; 3. Bias in classification of interventions; 4. Bias due to deviations from intended interventions; 5. Bias due to missing data; 6. Bias in measurement of outcomes; 7. Bias in selection of the reported result. For each domain, we could rate it as one of the following: Low risk of bias, Moderate risk of bias, Serious risk of bias, Critical risk of bias and No information, as well as the overall risk of bias.

Data Analysis

The characteristics of the study, patient and intervention will be summarized descriptively. We will make a narrative review for some comparisons if insufficient data is provided. Network plot will be drawn to descript the available interventions. The size of the node reflects the number of patients in each intervention. The breadth of the edge shows the number of comparisons. For efficacy outcomes, pair-wise and network meta-analysis will be performed for data from RCT. For safety outcomes, pair-wise and network meta-analysis will be performed for data from RCT and NRCT, separately.

Pairwise Meta-Analyses

We will perform traditional pair-wise meta-analyses through the random-effect model for every direct comparison. In some subgroups, we will also perform pair-wise meta-analyses if network meta-analyses could not be performed. The heterogeneity will be assessed by I-square and tau-square (23).

Assessment of the Transitivity Assumption

The potential baseline effect modifiers will be assessed to confirm they are similar among different comparisons before we perform network meta-analyses (16). If any difference is found, we will conduct meta-regression to explore the influence on the results.

Network Meta-Analyses

Random-effect network meta-analyses under the frequentist framework will be performed to combine both direct and indirect comparisons (29, 30). The heterogeneity parameter is assumed the same for each network (31, 32).

Assessment of Inconsistency

Bucher method as a local method and design-by-treatment interaction model as a global method will be used (33, 34). If any inconsistency is found, the node-splitting method will be used to explore the origin of the inconsistency (35).

Exploring Sources of Heterogeneity or Inconsistency with Subgroup Analyses and Meta-Regression

For two primary outcomes, subgroup analyses and meta-regression will be performed to assess the influence of the potential effect modifiers. Subgroup analyses will be presented as follows: 1. Gestational age by trimester; 2. The parity situation; 3. For postpartum outcomes, a subgroup about the duration of follow-up will be set for three months, one year and three years; 4. Patients with radicular pain versus without. Meta-regression will be performed as follows: 1. Age; 2. Sample size; 3. Baseline pain intensity; 4. Baseline physical function; 5. Percentage of the smoker.

[Sensitivity Analyses

For two primary outcomes, sensitivity analyses will be performed as follows: 1. Only studies with a low risk of bias; 2. Studies without a non-active comparison group; 3. Studies without receiving commercial funding; 4. Studies without unpublished data.

Publication Bias

Comparison-adjusted funnel plot will be used to test the publication bias if the number of included studies is larger than 10 (25). As above describes, meta-regression about sample size and effect estimates will be performed to detect the small-study effect (36).

Statistical Software

All analyses will be performed in Stata (StataCorp 2017. Stata Statistical Software: Release 15.1. College Station, TX: StataCorp LP).

ETHICS AND DISSEMINATION

This research does not require ethics approval because it uses data from literature. We will publish the study in a peer-reviewed journal after completing it.

References

1. Carvalho M, Lima LC, de Lira Terceiro CA, Pinto DRL, Silva MN, Cozer GA, et al. [Low back pain during pregnancy]. Revista brasileira de anestesiologia. 2017;67(3):266-70.

2. Katonis P, Kampouroglou A, Aggelopoulos A, Kakavelakis K, Lykoudis S, Makrigiannakis A, et al. Pregnancy-related low back pain. Hippokratia. 2011;15(3):205-10.

3. Sencan S, Ozcan-Eksi EE, Cuce I, Guzel S, Erdem B. Pregnancy-related low back pain in women in Turkey: Prevalence and risk factors. Annals of physical and rehabilitation medicine. 2018;61(1):33-7.

4. Ritchie JR. Orthopedic considerations during pregnancy. Clinical obstetrics and gynecology. 2003;46(2):456-66.

5. Noren L, Ostgaard S, Johansson G, Ostgaard HC. Lumbar back and posterior pelvic pain during pregnancy: a 3-year follow-up. European spine journal : official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society. 2002;11(3):267-71.

6. Padua L, Caliandro P, Aprile I, Pazzaglia C, Padua R, Calistri A, et al. Back pain in pregnancy: 1-year follow-up of untreated cases. European spine journal : official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society. 2005;14(2):151-4.

7. Ostgaard HC, Andersson GB. Previous back pain and risk of developing back pain in a future pregnancy. Spine. 1991;16(4):432-6.

8. Casagrande D, Gugala Z, Clark SM, Lindsey RW. Low Back Pain and Pelvic Girdle Pain in Pregnancy. The Journal of the American Academy of Orthopaedic Surgeons. 2015;23(9):539-49.

9. Sinclair. M, Close. C, McCullough. JE, Hughes. C, Liddle. SD. How do women manage pregnancy-related low back and/or pelvic pain? Descriptive findings from an online survey. Evidence Based Midwifery. 2014;12(3):76-82.

10. Liddle SD, Pennick V. Interventions for preventing and treating low-back and pelvic pain during pregnancy. The Cochrane database of systematic reviews. 2015(9):Cd001139.

11. Kinser PA, Pauli J, Jallo N, Shall M, Karst K, Hoekstra M, et al. Physical Activity and Yoga-Based Approaches for Pregnancy-Related Low Back and Pelvic Pain. Journal of obstetric, gynecologic, and neonatal nursing : JOGNN. 2017;46(3):334-46.

12. Sehmbi H, D'Souza R, Bhatia A. Low Back Pain in Pregnancy: Investigations, Management, and Role of Neuraxial Analgesia and Anaesthesia: A Systematic Review. Gynecologic and obstetric investigation. 2017;82(5):417-36.

13. Stuber KJ, Smith DL. Chiropractic treatment of pregnancy-related low back pain: a systematic review of the evidence. Journal of manipulative and physiological therapeutics. 2008;31(6):447-54.

14. Stuge B, Hilde G, Vøllestad N. Physical therapy for pregnancy-related low back and pelvic pain: a systematic review. Acta Obstetricia et Gynecologica Scandinavica. 2003;82(11):983-90.

15. Yao X, Li C, Ge X, Wei J, Luo J, Tian F. Effect of acupuncture on pregnancy related low back pain and pelvic pain: A systematic review and meta-analysis. International Journal of Clinical and Experimental Medicine. 2017;10(4):5903-12.

16. Dias S, Welton NJ, Sutton AJ, Ades AE. Evidence synthesis for decision making 1: introduction. Medical decision making : an international journal of the Society for Medical Decision Making. 2013;33(5):597-606.

17. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Systematic Reviews. 2015;4(1):1.

18. Clement RC, Welander A, Stowell C, Cha TD, Chen JL, Davies M, et al. A proposed set of metrics for standardized outcome reporting in the management of low back pain. Acta orthopaedica. 2015;86(5):523-33.

19. Chapman JR, Norvell DC, Hermsmeyer JT, Bransford RJ, DeVine J, McGirt MJ, et al. Evaluating common outcomes for measuring treatment success for chronic low back pain. Spine. 2011;36(21 Suppl):S54-68.

20. Jensen MP, Mardekian J, Lakshminarayanan M, Boye ME. Validity of 24-h recall ratings of pain severity: biasing effects of "Peak" and "End" pain. Pain. 2008;137(2):422-7.

21. Dakin H A, L, Burns, R, Yang, Y. . Review and critical appraisal of studies mapping from quality of life or clinical measures to EQ-5D: an online database and application of the MAPS statement. Health and Quality of Life Outcomes. 16:31 HERC database of mapping studies, Version 6.0 (Last updated: 17th Jan 2017). Available at: http://www.herc.ox.ac.uk/downloads/herc-database-of-mapping-studies. 2018.

22. Caddy C, Amit BH, McCloud TL, Rendell JM, Furukawa TA, McShane R, et al. Ketamine and other glutamate receptor modulators for depression in adults. The Cochrane database of systematic reviews. 2015(9):Cd011612.

23. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.handbook.cochrane.org.

24. Furlan AD, Malmivaara A, Chou R, Maher CG, Deyo RA, Schoene M, et al. 2015 Updated Method Guideline for Systematic Reviews in the Cochrane Back and Neck Group. Spine. 2015;40(21):1660-73.

25. Salanti G, Ades AE, Ioannidis JP. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. Journal of clinical epidemiology. 2011;64(2):163-71.

26. van Tulder MW, Suttorp M, Morton S, Bouter LM, Shekelle P. Empirical evidence of an association between internal validity and effect size in randomized controlled trials of low-back pain. Spine. 2009;34(16):1685-92.

27. van Tulder MW BG CR, Corbin T, Deyo RA, Furlan AD, Malmivaara A, Marin T, Peul W, Schoene ML, Maher CG, Weiner BK. Cochrane Back and Neck Group. About The Cochrane Collaboration (Cochrane Review Groups (CRGs)) 2014, Issue 2. Art. No.: BACK.

28. Sterne JA, Hernan MA, Reeves BC, Savovic J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ (Clinical research ed). 2016;355:i4919.

29. Dias S, Sutton AJ, Ades AE, Welton NJ. Evidence synthesis for decision making 2: a generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. Medical decision making : an international journal of the Society for Medical Decision Making. 2013;33(5):607-17.

30. IR W. Multivariate random-effects meta-analysis. Stata Journal. 2009;9(1):40-56.

31. Dias S, Sutton AJ, Welton NJ, Ades AE. Evidence synthesis for decision making 3: heterogeneity--subgroups, meta-regression, bias, and bias-adjustment. Medical decision making : an international journal of the Society for Medical Decision Making. 2013;33(5):618-40.

32. Bighelli I, Salanti G, Reitmeir C, Wallis S, Barbui C, Furukawa TA. Psychological interventions for positive symptoms in schizophrenia: protocol for a network meta-analysis of randomised controlled trials. 2018;8(3):e019280.

33. Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. Journal of clinical epidemiology. 1997;50(6):683-91.

34. Higgins JP, Jackson D, Barrett JK, Lu G, Ades AE, White IR. Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. Research synthesis methods. 2012;3(2):98-110.

35. Dias S, Welton NJ, Sutton AJ, Caldwell DM, Lu G, Ades AE. Evidence synthesis for decision making 4: inconsistency in networks of evidence based on randomized controlled trials. Medical decision making : an international journal of the Society for Medical Decision Making. 2013;33(5):641-56.

36. Chaimani A, Salanti G. Using network meta-analysis to evaluate the existence of small-study effects in a network of interventions. Research synthesis methods. 2012;3(2):161-76.

Search Strategy

Database: Ovid MEDLINE(R) Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R)

1 exp Low Back Pain/ (18896)

2 exp Backache/ (34916)

3 "Low back pain".mp. (29875)

4 Back Pain.mp. (54017)

5 Backpain.mp. (69)

6 Backache.mp. (3538)

7 Back ache.mp. (98)

8 (lumbar adj5 pain).ti,ab. (4860)

9 Lumbar pain.mp. (1345)

10 Spinal pain.mp. (1264)

11 Lumbago.mp. (1293)

12 Lower back pain.mp. (2055)

13 Dorsalgia.mp. (86)

14 Vertebral pain.mp. (103)

15 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 (58834)

16 exp Pregnancy/ (830174)

17 exp Pregnancy Complications/ (393355)

18 exp Maternal Health Services/ (42806)

19 pregnant women/ (6602)

20 (pregnan* or postpartum).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol Supplementaryary concept word, rare disease Supplementaryary concept word, unique identifier, synonyms] (940534)

21 Perinatal Care/ (3824)

22 exp Postpartum Period/ (57970)

23 16 or 17 or 18 or 19 or 20 or 21 or 22 (993450)

24 randomized controlled trial.pt. (457341)

25 controlled clinical trial.pt. (92294)

26 randomized.ab,ti. (438936)

27 placebo.ab,ti. (192887)

28 drug therapy.fs. (2005857)

29 randomly.ab,ti. (288865)

30 trial.ab,ti. (497372)

31 groups.ab,ti. (1805786)

32 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 (4227862)

33 (animals not (humans and animals)).sh. (4408192)

34 32 not 33 (3657644)

35 Non-Randomized Controlled Trials as Topic/ (321)

36 ((nonrandom* or non-random* or quasi-random* or quasi-experiment*) adj (stud* or trial*)).tw. (11747)

37 (non-RCT or non-RCTs or nRCT or nRCTs).tw. (626)

38 exp cohort studies/ (1726989)

39 (cohort stud* or Follow-Up Stud* or Longitudinal Stud* or Prospective Stud* or Retrospective Stud*).tw. (518582)

40 exp case-control studies/ (906482)

41 case-control* stud*.tw. (90488)

42 35 or 36 or 37 or 38 or 39 or 40 or 41 (2110451)

43 34 or 42 (5084336)

44 15 and 23 and 43 (668)

Database: Embase

1 dorsalgia.mp. (138)

2 back pain.mp. (77876)

3 exp BACKACHE/ (94362)

4 (lumbar adj pain).mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word] (2070)

5 coccyx.mp. (1088)

6 coccydynia.mp. (172)

7 sciatica.mp. (5587)

8 exp ISCHIALGIA/ (5335)

9 spondylosis.mp. (8719)

10 lumbago.mp. (1701)

11 exp Low back pain/ (49650)

12 back disorder$.mp. (709)

13 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 (119518)

14 exp PREGNANCY/ (690864)

15 exp PREGNANCY DISORDER/ (512296)

16 pregnan*.mp. (928401)

17 Postpartum.mp. (63319)

18 14 or 15 or 16 or 17 (1109487)

19 Clinical Article/ (1871963)

20 exp Clinical Study/ (8484461)

21 Clinical Trial/ (969210)

22 Controlled Study/ (5896797)

23 Randomized Controlled Trial/ (497763)

24 Major Clinical Study/ (3075274)

25 Double Blind Procedure/ (148714)

26 Multicenter Study/ (182338)

27 Single Blind Procedure/ (31030)

28 Phase 3 Clinical Trial/ (33517)

29 Phase 4 Clinical Trial/ (2932)

30 crossover procedure/ (55089)

31 placebo/ (323401)

32 allocat$.mp. (140211)

33 assign$.mp. (336185)

34 blind$.mp. (436000)

35 (clinic$ adj25 (study or trial)).mp. (5546771)

36 compar$.mp. (7018177)

37 control$.mp. (8965932)

38 cross?over.mp. (87010)

39 factorial$.mp. (61125)

40 follow?up.mp. (41983)

41 placebo$.mp. (417193)

42 prospectiv$.mp. (986175)

43 random$.mp. (1492799)

44 ((singl$ or doubl$ or trebl$ or tripl$) adj25 (blind$ or mask$)).mp. (280252)

45 trial.mp. (1872782)

46 (versus or vs).mp. (1811258)

47 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46 (17864569)

48 exp animals/ or exp invertebrate/ or animal experiment/ or animal model/ or animal tissue/ or animal cell/ or nonhuman/ (25892717)

49 human/ or normal human/ or human cell/ (19576403)

50 48 and 49 (19528170)

51 48 not 50 (6364547)

52 47 not 51 (14341811)

53 cohort analysis/ (361958)

54 exp case control study/ (142056)

55 case-control* stud*.mp. (180962)

56 follow up/ (1276156)

57 exp longitudinal study/ (111132)

58 prospective study/ (440403)

59 retrospective study/ (636210)

60 quasi experimental study/ (4474)

61 (cohort stud* or Follow-Up Stud* or Longitudinal Stud* or Prospective Stud* or Retrospective Stud*).mp. (1478425)

62 "controlled clinical trial (topic)"/ (9400)

63 (non-RCT or non-RCTs or nRCT or nRCTs).mp. (928)

64 ((nonrandom* or non-random* or quasi-random* or quasi-experiment*) adj (stud* or trial*)).mp. (17486)

65 53 or 54 or 55 or 56 or 57 or 58 or 59 or 60 or 61 or 62 or 63 or 64 (2682771)

66 52 or 65 (14593960)

67 13 and 18 and 66 (2603)

Database: The Cochrane Library

#1¬MeSH descriptor Pregnancy explode all trees
#2¬MeSH descriptor Pregnancy Complications explode all trees

#3¬MeSH descriptor Maternal Health Services explode all trees

#4¬MeSH descriptor Perinatal Care explode all trees

#5¬MeSH descriptor Postpartum Period explode all trees

#6¬pregnan* in All Fields in all products

#7¬postpartum in All Fields in all products

#8-#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7

#9 MeSH descriptor Back Pain explode all trees

#10 dorsalgia

#11 backache

#12 MeSH descriptor Low Back Pain explode all trees

#13 lumbar next pain OR coccyx OR coccydynia OR sciatica OR spondylosis

#14 MeSH descriptor Sciatica explode all trees

#15 MeSH descriptor Spine explode all trees

#16 MeSH descriptor Spinal Diseases explode all trees

#17 lumbago OR discitis OR disc near degeneration OR disc near prolapse OR disc near herniation

#18 spinal fusion

#19 spinal neoplasms

#20 facet near joints

#21 MeSH descriptor Intervertebral Disk explode all trees

#22 postlaminectomy

#23 arachnoiditis

#24 failed near back

#25 MeSH descriptor Cauda Equina explode all trees

#26 lumbar near vertebra*

#27 spinal near stenosis

#28 slipped near (disc* or disk*)

#29 degenerat* near (disc* or disk*)

#30 stenosis near (spine or root or spinal)

#31 displace* near (disc* or disk*)

#32 prolap* near (disc* or disk*)

#33 #9 or #10 or #11 or #12 or #13 or #14 or #15 or #16 or #17 or #18 or #19 or #20 or #21 or #22 or #23 or #24 or #25 or #26 or #27 or #28 or #29 or #30 or #31 or #32

#34 #8 AND #33

Database: AMED (Allied and Complementary Medicine)

1 exp Low back pain/ or low back pain.mp. (5447)

2 back pain.mp. or exp Backache/ (7223)

3 exp Neck pain/ or neck pain.mp. (1440)

4 (low back pain or back pain or neck pain or backache or lumbago or neck ache or spin* pain or knee pain or hip pain).mp. (9140)

5 1 or 2 or 3 or 4 (9184)

6 pregnancy/ (1255)

7 pregnancy complications/ (636)

8 postpartum.mp. (190)

9 pregnan*.mp. (2123)

10 6 or 7 or 8 or 9 (2177)

11 exp Randomized controlled trials/ or randomized controlled trial.mp. (4289)

12 randomized controlled trial.pt. (4074)

13 exp Random allocation/ or random allocation.mp. (354)

14 exp Placebos/ or placebo.mp. (3049)

15 (random* adj3 trial).ab,ti. (5633)

16 Random*.ab,ti. (16750)

17 11 or 12 or 13 or 14 or 15 or 16 (18599)

18 case control studies/ (143)

19 cohort studies/ (982)

20 follow up studies/ (1403)

21 longitudinal studies/ (579)

22 prospective studies/ (1047)

23 retrospective studies/ (659)

24 case-control* stud*.mp. (644)

25 (cohort stud* or Follow-Up Stud* or Longitudinal Stud* or Prospective Stud* or Retrospective Stud*).mp. (8918)

26 (non-RCT or non-RCTs or nRCT or nRCTs).mp. (15)

27 ((nonrandom* or non-random* or quasi-random* or quasi-experiment*) adj (stud* or trial*)).mp. (209)

28 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 (9693)

29 17 or 28 (27044)

30 5 and 10 and 29 (31)

Database: CINAHL

S57 S17 AND S24 AND S56 442

S56 S48 OR S55 1,255,387

S55 S49 OR S50 OR S51 OR S52 OR S53 OR S54 377,256

S54 ((nonrandom* or non-random* or quasi-random* or quasi-experiment*) AND (stud* or trial*)) 15,734

S53 (MH "Quasi-Experimental Studies+") 10,460

S52 cohort stud* or Follow-Up Stud* or Longitudinal Stud* or Prospective Stud* or Retrospective Stud* 276,046

S51 case-control* stud* 39,851

S50 (MH "Retrospective Design") OR (MH "Retrospective Panel Studies") 104,735

S49 (MH "Prospective Studies+") OR (MH "Case Control Studies+") 247,700

S48 S46 not S47 1,242,625

S47 (MH "Animals+") 37,347

S46 S25 OR S26 OR S27 OR S28 OR S29 OR S30 OR S31 OR S32 OR S33 OR S34 OR S35 OR S36 OR S37 OR S38 OR S39 OR S40 OR S41 OR S42 OR S43 OR S44 OR S45 1,255,981

S45 volunteer* 28,994

S44 prospectiv* 250,036

S43 control* 670,584

S42 followup stud* 396

S41 follow-up stud* 13,597

S40 (MH "Prospective Studies+") 215,887

S39 (MH "Evaluation Research+") 42,071

S38 (MH "Comparative Studies") 103,577

S37 latin square 142

S36 (MH "Study Design+") 694,900

S35 (MH "Random Sample+") 68,215

S34 random* 204,969

S33 placebo* 33,851

S32 (MH "Placebos") 8,341

S31 (MH "Placebo Effect") 1,216

S30 triple-blind 139

S29 single blind 9,042

S28 double blind 29,448

S27 clinical W3 trial 124,800

S26 randomi?ed controlled trial* 77,785

S25 (MH "Clinical Trials+") 156,485

S24 S18 OR S19 OR S20 OR S21 OR S22 OR S23 141,421

S23 (MH "Pregnancy Trimesters+") 4,889

S22 (MH "Pregnancy, Multiple+") 1,682

S21 pregnan* 127,526

S20 Postpartum 13,180

S19 (MH "Pregnancy Complications+") 47,621

S18 (MH "Pregnancy+") 117,854

S17 S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13 OR S14 OR S15 OR S16 29,571

S16 lumbago 38

S15 (MH "Spondylolysis") 272

S14 (MH "Spondylolisthesis") 587

S13 lumbar N2 vertebrae 7,339

S12 (MH "Lumbar Vertebrae") 7,227

S11 back disorder* 517

S10 coccydynia 39

S9 coccyx 161

S8 sciatica 1,057

S7 (MH "Sciatica") 789

S6 (MH "Coccyx") 111

S5 lumbar N5 pain 1,283

S4 lumbar W1 pain 354

S3 backache or back pain 23,197

S2 (MH "Back Pain+") 19,102

S1 dorsalgia 8

Database: PEDro

Abstract & Title: pregnan* OR postpartum

AND

Problem: pain

AND

Body part: Lumbar spine, sacroiliac joint or pelvis

AND

Method: Clinical Trial

Database: PsycINFO

1 back pain/ (3549)

2 dorsalgia.mp. (6)

3 backache.mp. (134)

4 (lumbar adj pain).mp. (53)

5 (low adj back adj pain).mp. (3243)

6 sciatica.mp. (146)

7 lumbago.mp. (35)

8 spinal nerves/ (2535)

9 lumbar spinal cord/ (625)

10 ((disc or disk) adj degenerat*).mp. (36)

11 ((disc or disk) adj prolapse*).mp. (19)

12 ((disc or disk) adj herniat*).mp. (137)

13 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 (7646)

14 Pregnancy Outcomes/ (956)

15 exp Pregnancy/ (22930)

16 Obstetrical Complications/ (1246)

17 Prenatal Care/ (1659)

18 pregnan*.mp. (45293)

19 postpartum.mp. (11662)

20 14 or 15 or 16 or 17 or 18 or 19 (52461)

21 clinical trials/ (10856)

22 Randomi?ed controlled trial*.mp. (25501)

23 control*.mp. (654443)

24 random*.mp. (177536)

25 exp Treatment/ (709214)

26 21 or 22 or 23 or 24 or 25 (1330923)

27 cohort analysis/ (1248)

28 followup studies/ (12359)

29 longitudinal studies/ (15441)

30 prospective studies/ (498)

31 retrospective studies/ (389)

32 quasi experimental methods/ (144)

33 case-control* stud*.mp. (6848)

34 (cohort stud* or Follow-Up Stud* or Longitudinal Stud* or Prospective Stud* or Retrospective Stud*).mp. (100060)

35 (non-RCT or non-RCTs or nRCT or nRCTs).mp. (70)

36 ((nonrandom* or non-random* or quasi-random* or quasi-experiment*) adj (stud* or trial*)).mp. (2891)

37 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 (119499)

38 26 or 37 (1402501)

39 13 and 20 and 38 (39)

Database: Clinicaltrials.gov

pain | Studies With Results | pregnan* OR postpartum | Studies with Female Participants

Supplementary Table 1. Magnitude of Effects on pain intensity and physical function.

Pain Intensity

Rating Scale	Slight/Small	Moderate	Large/Substantial
	If the studies use the same rating scale		
VAS (0-100)	5-10	10-20	>20
NRS (0-10)	0.5-1	1-2	>2
	If the studies use the different rating scales		
SMD	0.2-0.5	0.5-0.8	>0.8

Physical Function

Rating Scale	Slight/Small	Moderate	Large/Substantial
	If the studies use the same rating scale		
ODI (0-100)	5-10	10-20	>20
RMDQ (0-24)	1-2	2-5	>5
	If the studies use the different rating scales		
SMD	0.2-0.5	0.5-0.8	>0.8

VAS: Visual Analog Scale, NRS: Numeric Rating Scale, ODI: Oswestry Disability Index, RMDQ: Roland Morris Disability Questionnaire, SMD: Standardized Mean Difference.

Supplementary Table 2. Baselines of Primary Outcomes in Included Studies [a]

During pregnancy					
Study	Comparisons	Pain		Physical function	
		Methods	Baselines	Methods	Baselines
Abu 2017	Exercise + Paracetamol	VAS	4 (4)	Oswestry Disability Index	26 (16)
	Paracetamol		5 (4)		24 (14)
Akmeses 2014	Progressive muscle relaxation	VAS	7.78 (1.61)	36-Item Short Form Survey	76.21 (9.1)
	Control group		7.69 (1.75)		67.87 (9.84)
Mohseni Bandpei	Exercise	VAS	5.31 (1.18)	Oswestry Disability Index	40.7 (14.22)
et al 2010	Control group		5.73 (0.78)		41.5 (13.29)
Gil et al 2011	Exercise	VAS	5.2	Roland Morris Disability Questionnaire	7.1
	Control group		5.8		9.5
Gundermann 2013	Spinal manipulative treatment	VAS	6.1 (1.4)	Roland Morris Disability Questionnaire	6.7 (3.4)
	Control group		5.8 (1.4)		6.2 (3.3)
Hensel 2015	Spinal manipulative treatment	VAS	5.07 (2.39)	Roland Morris Disability Questionnaire	6.7 (4.97)
	Placebo		4.71 (2.34)		5.9 (4.68)
	Control group		4.78 (2.41)		6.55 (5.09)
Kalinowski 2017	Kinesio taping	VAS	4.94 (1.74)	Roland Morris Disability Questionnaire	5.66 (2.64)
	Placebo		4.96 (1.73)		5.4 (2.21)
Kalus 2008	BellyBra	VAS	6.1 (2.2)	Likert scale	6.5 (2.3)
	Tubigrip		6 (2)		6.4 (1.7)
Kaplan 2016	Kinesio taping + Paracetamol	VAS	7.57 (1.49)	Roland Morris Disability Questionnaire	13.42 (3.82)

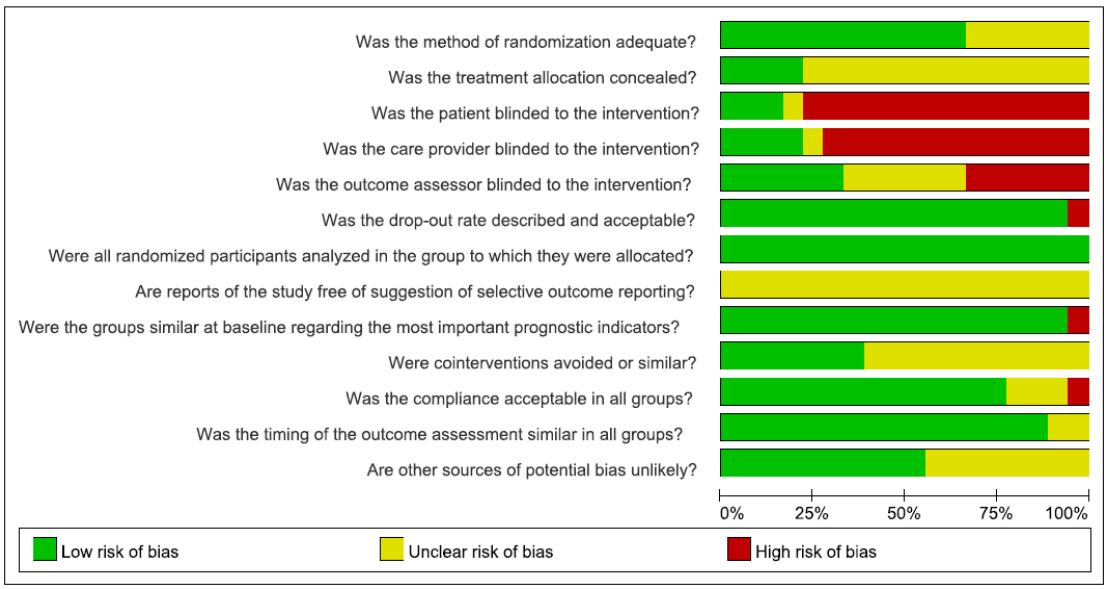
	Paracetamol		7.27 (1.13)		15.03 (3.56)
Keskin 2012	TENS	VAS	7 (0.74)	Roland Morris Disability Questionnaire	15 (3.70)
	Paracetamol		6 (0.74)		14 (2.22)
	Exercise		7 (0.74)		15(2.96)
	Control group		6 (0.74)		14 (0.74)
Licciardone 2010	Spinal manipulative treatment	NRS	4.9 (2.1)	Roland Morris Disability Questionnaire	8.4 (4.7)
	Placebo		4.8 (2.3)		8.1 (5.3)
	Control group		4.9 (2.3)		6.6 (4.5)
Pekçetin et al	Tel	VAS	6.66 (2.16)	Oswestry Disability Index	39.53 (14.04)
2019	Fac		5.15 (1.68)		42.96 (12.84)
Rohrich 2014	Spinal manipulative treatment	VAS	6.1 (0.8)	Roland Morris Disability Questionnaire	7.4 (3.8)
	Control group		5.7 (1.7)		6 (3.3)
Sedaghati 2007	Exercise	Quebec	20.43 (7.25)	N/A	N/A
	Control group		21.88 (7.06)		
Shirazi 2017	Rose oil	VAS	5.86	Roland Morris Disability Questionnaire	9.97
	Placebo		5.18		8.15
	Control group		5.05		6
Yan et al 2014	Exercise	BPI-T	7.67 (4.68)	Family Exercise Support Attitude	17.73 (11.99)
	Control group		9.61 (4.07)	Questionnaire	24.25 (12.3)
Postpartum					
Akbarzadeh 2014	Cupping	VAS	7.8 (2.7)	N/A	N/A
	Control group		7.6 (2.7)		
Belz 2014	Spinal manipulative treatment	VAS	6.95	Pelvic Girdle Pain Questionnaire	N/A

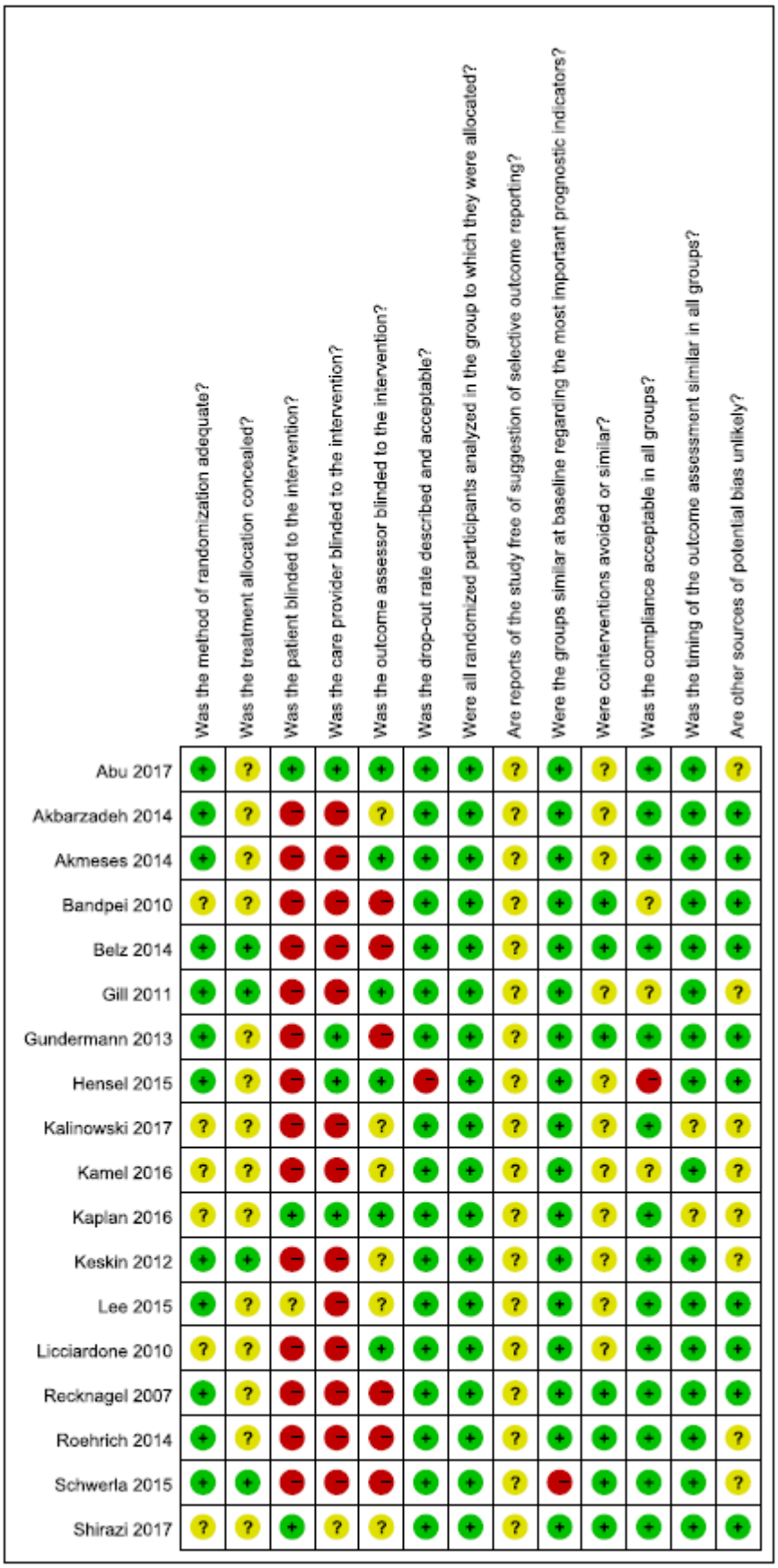
	Control group		6.41		
Kamel 2016	Spinal manipulative treatment	VAS	7.2 (1.08)	Oswestry Disability Index	57.14 (9.96)
	Placebo		7.26 (0.96)		56.72 (7.63)
	Control group		7.53 (1.06)		53.63 (11.58)
Lee 2015	Spinal manipulative treatment	VAS	5.02 (1.97)	N/A	N/A
	Control group		4.7 (1.8)		
Mohamed 2018	Kinesio taping + Exercise	VAS	6.95 (1.23)	Back Pain Function Scale	21 (4.82)
	Exercise		7.65 (1.08)		19.7 (3.62)
Recknagel 2007	Spinal manipulative treatment	VAS	6.83 (1.41)	Pelvic Girdle Pain Questionnaire	28.1 (12.2)
	Control group		5.92 (0.83)		28.5 (9.4)
Schwerla 2015	Spinal manipulative treatment	VAS	7.3 (0.9)	Oswestry Disability Index	16.8 (6.7)
	Control group		7 (1)		22.1 (7.2)

Pain and physical function are presented through mean (stand deviance).

[a] VAS: Visual Analog Scale, NRS: Numeric Rating Scale, Quebec: Quebec questionnaire, BPI-T: Brief Pain Inventory–Short Form Taiwanese Version. TENS: transcutaneous electrical nerve stimulation, Tel: telephone-supported ergonomic education, Fac: face-to-face ergonomic education, N/A: not applicable.

Supplementary Figure 1. Risk of bias




Was the method of randomization adequate?
Was the treatment allocation concealed?
Was the patient blinded to the intervention?
Was the care provider blinded to the intervention?
Was the outcome assessor blinded to the intervention?
Was the drop-out rate described and acceptable?
Were all randomized participants analyzed in the group to which they were allocated?
Are reports of the study free of suggestion of selective outcome reporting?
Were the groups similar at baseline regarding the most important prognostic indicators?
Were cointerventions avoided or similar?
Was the compliance acceptable in all groups?
Was the timing of the outcome assessment similar in all groups?
Are other sources of potential bias unlikely?
Abu 2017
Akbarzadeh 2014
Akmeses 2014
Bandpei 2010
Belz 2014
Gill 2011
Gundermann 2013
Hensel 2015
Kalinowski 2017
Kamel 2016
Kaplan 2016
Keskin 2012
Lee 2015
Licciardone 2010
Recknagel 2007
Roehrich 2014
Schwerla 2015
Shirazi 2017

Supplementary Table 3. GRADE Results

3.1 Pain intensity for LBP during pregnancy

Comparison	Number of Studies	Within-Study Bias	Across-Studies Bias	Indirectness	Imprecision	Heterogeneity	Incoherence	Confidence rating
Con:Exe	3	No concerns	Undetected	No concerns	Some concerns	No concerns	No concerns	Moderate
Con:PMR	1	No concerns	Undetected	No concerns	No concerns	Some concerns	No concerns	Moderate
Con:Par	1	No concerns	Undetected	No concerns	Some concerns	No concerns	Some concerns	Moderate
Con:Pla	3	No concerns	Undetected	No concerns	Some concerns	No concerns	No concerns	Moderate
Con:Ros	1	No concerns	Undetected	No concerns	Some concerns	No concerns	Some concerns	Moderate
Con:SMT	4	No concerns	Undetected	No concerns	Some concerns	No concerns	No concerns	Moderate
Con:TENS	1	No concerns	Undetected	No concerns	Some concerns	No concerns	Some concerns	Moderate
Exe:Par	1	No concerns	Undetected	No concerns	Some concerns	Some concerns	Major concerns	Low
Exe:TENS	1	No concerns	Undetected	No concerns	Some concerns	Some concerns	Major concerns	Low
Exe+Par:Par	1	No concerns	Undetected	No concerns	Some concerns	Some concerns	No concerns	Moderate
Kin:Pla	1	Major concerns	Undetected	No concerns	Some concerns	No concerns	No concerns	Low

Kin+Par:Par	1	Major concerns	Undetected	No concerns	Some concerns	No concerns	No concerns	Low
Par:TENS	1	No concerns	Undetected	No concerns	Some concerns	Some concerns	No concerns	Moderate
Pla:Ros	1	No concerns	Undetected	No concerns	Some concerns	No concerns	No concerns	Moderate
Pla:SMT	2	No concerns	Undetected	No concerns	No concerns	Major concerns	Some concerns	Low
Con:Exe+Par	0	No concerns	Undetected	No concerns	Some concerns	No concerns	Some concerns	Moderate
Con:Kin	0	No concerns	Undetected	No concerns	No concerns	Some concerns	Some concerns	Moderate
Con:Kin+Par	0	No concerns	Undetected	No concerns	Some concerns	No concerns	Some concerns	Moderate
Exe:Exe+Par	0	No concerns	Undetected	No concerns	Some concerns	Some concerns	Some concerns	Low
Exe:Kin	0	No concerns	Undetected	No concerns	Some concerns	No concerns	Some concerns	Moderate
Exe:Kin+Par	0	No concerns	Undetected	No concerns	Some concerns	Some concerns	Some concerns	Low
Exe:PMR	0	No concerns	Undetected	No concerns	Some concerns	No concerns	Some concerns	Moderate
Exe:Pla	0	No concerns	Undetected	No concerns	Some concerns	Some concerns	Some concerns	Low
Exe:Ros	0	No concerns	Undetected	No concerns	Some concerns	Some concerns	Some concerns	Low
Exe:SMT	0	No concerns	Undetected	No concerns	Some concerns	Some concerns	Some concerns	Low
Exe+Par:Kin	0	No concerns	Undetected	No concerns	Major concerns	No concerns	Some concerns	Low

Exe+Par:Kin+Par	0	No concerns	Undetected	No concerns	Major concerns	No concerns	Some concerns	Low
Exe+Par:PMR	0	No concerns	Undetected	No concerns	Some concerns	Some concerns	Some concerns	Low
Exe+Par:Pla	0	No concerns	Undetected	No concerns	Some concerns	Some concerns	Some concerns	Low
Exe+Par:Ros	0	No concerns	Undetected	No concerns	Major concerns	No concerns	Some concerns	Low
Exe+Par:SMT	0	No concerns	Undetected	No concerns	Some concerns	Some concerns	Some concerns	Low
Exe+Par:TENS	0	No concerns	Undetected	No concerns	Major concerns	No concerns	Some concerns	Low
Kin:Kin+Par	0	No concerns	Undetected	No concerns	Major concerns	No concerns	Some concerns	Low
Kin:PMR	0	No concerns	Undetected	No concerns	Major concerns	No concerns	Some concerns	Low
Kin:Par	0	No concerns	Undetected	No concerns	Some concerns	Some concerns	Some concerns	Low
Kin:Ros	0	No concerns	Undetected	No concerns	Some concerns	Some concerns	Some concerns	Low
Kin:SMT	0	No concerns	Undetected	No concerns	Some concerns	No concerns	Some concerns	Moderate
Kin:TENS	0	No concerns	Undetected	No concerns	Some concerns	Some concerns	Some concerns	Low
Kin+Par:PMR	0	No concerns	Undetected	No concerns	Major concerns	No concerns	Some concerns	Low
Kin+Par:Pla	0	No concerns	Undetected	No concerns	Some concerns	No concerns	Some concerns	Moderate
Kin+Par:Ros	0	No concerns	Undetected	No concerns	Some concerns	Some concerns	Some concerns	Low

Kin+Par:SMT	0	No concerns	Undetected	No concerns	Some concerns	No concerns	Some concerns	Moderate
Kin+Par:TENS	0	No concerns	Undetected	No concerns	Major concerns	No concerns	Some concerns	Low
PMR:Par	0	No concerns	Undetected	No concerns	Some concerns	No concerns	Some concerns	Moderate
PMR:Pla	0	No concerns	Undetected	No concerns	Some concerns	No concerns	Some concerns	Moderate
PMR:Ros	0	No concerns	Undetected	No concerns	Some concerns	No concerns	Some concerns	Moderate
PMR:SMT	0	No concerns	Undetected	No concerns	Some concerns	No concerns	Some concerns	Moderate
PMR:TENS	0	No concerns	Undetected	No concerns	Some concerns	Some concerns	Some concerns	Low
Par:Pla	0	No concerns	Undetected	No concerns	Some concerns	Some concerns	Some concerns	Low
Par:Ros	0	No concerns	Undetected	No concerns	Major concerns	No concerns	Some concerns	Low
Par:SMT	0	No concerns	Undetected	No concerns	Some concerns	Some concerns	Some concerns	Low
Pla:TENS	0	No concerns	Undetected	No concerns	Some concerns	No concerns	Some concerns	Moderate
Ros:SMT	0	No concerns	Undetected	No concerns	Some concerns	Some concerns	Some concerns	Low
Ros:TENS	0	No concerns	Undetected	No concerns	Some concerns	Some concerns	Some concerns	Low
SMT:TENS	0	No concerns	Undetected	No concerns	Some concerns	No concerns	Some concerns	Moderate

3.2 Physical function for LBP during pregnancy

Comparison	Number of studies	Within-study bias	Across-studies bias	Indirectness	Imprecision	Heterogeneity	Incoherence	Confidence rating
Con:Exe	2	No concerns	Undetected	No concerns	Some concerns	No concerns	No concerns	Moderate
Con:Par	1	No concerns	Undetected	No concerns	Some concerns	No concerns	No concerns	Moderate
Con:Pla	3	No concerns	Undetected	No concerns	No concerns	Some concerns	Some concerns	Moderate
Con:Ros	1	No concerns	Undetected	No concerns	Some concerns	No concerns	Some concerns	Moderate
Con:SMT	4	No concerns	Undetected	No concerns	No concerns	Some concerns	Some concerns	Moderate
Con:TENS	1	No concerns	Undetected	No concerns	Some concerns	No concerns	Some concerns	Moderate
Exe:Par	1	No concerns	Undetected	No concerns	No concerns	Major concerns	Major concerns	Low
Exe:TENS	1	No concerns	Undetected	No concerns	Some concerns	No concerns	Some concerns	Moderate
Kin:Pla	1	Major concerns	Undetected	No concerns	Some concerns	Some concerns	No concerns	Low
Kin+Par:Par	1	No concerns	Undetected	No concerns	Some concerns	No concerns	No concerns	Moderate
Par:TENS	1	No concerns	Undetected	No concerns	Some concerns	No concerns	No concerns	Moderate
Pla:Ros	1	No concerns	Undetected	No concerns	Some concerns	Some concerns	Some concerns	Low

Pla:SMT	2	No concerns	Undetected	No concerns	No concerns	Major concerns	Some concerns	Low
Con:Kin	0	No concerns	Undetected	No concerns	Some concerns	No concerns	Some concerns	Moderate
Con:Kin+Par	0	No concerns	Undetected	No concerns	Major concerns	No concerns	Some concerns	Low
Exe:Kin	0	No concerns	Undetected	No concerns	Major concerns	No concerns	Some concerns	Low
Exe:Kin+Par	0	No concerns	Undetected	No concerns	Some concerns	Some concerns	Some concerns	Low
Exe:Pla	0	No concerns	Undetected	No concerns	Some concerns	Some concerns	Some concerns	Low
Exe:Ros	0	No concerns	Undetected	No concerns	Major concerns	No concerns	Some concerns	Low
Exe:SMT	0	No concerns	Undetected	No concerns	Some concerns	Some concerns	Some concerns	Low
Kin:Kin+Par	0	No concerns	Undetected	No concerns	Some concerns	Some concerns	Some concerns	Low
Kin:Par	0	No concerns	Undetected	No concerns	Major concerns	No concerns	Some concerns	Low
Kin:Ros	0	No concerns	Undetected	No concerns	Major concerns	No concerns	Some concerns	Low
Kin:SMT	0	No concerns	Undetected	No concerns	Some concerns	Some concerns	Some concerns	Low
Kin:TENS	0	No concerns	Undetected	No concerns	Some concerns	No concerns	Some concerns	Moderate
Kin+Par:Pla	0	No concerns	Undetected	No concerns	Some concerns	Some concerns	Some concerns	Low
Kin+Par:Ros	0	No concerns	Undetected	No concerns	Some concerns	Some concerns	Some concerns	Low

Kin+Par:SMT	0	No concerns	Undetected	No concerns	Some concerns	Some concerns	Some concerns	Low
Kin+Par:TENS	0	No concerns	Undetected	No concerns	Some concerns	No concerns	Some concerns	Moderate
Par:Pla	0	No concerns	Undetected	No concerns	Some concerns	Some concerns	Some concerns	Low
Par:Ros	0	No concerns	Undetected	No concerns	Major concerns	No concerns	Some concerns	Low
Par:SMT	0	No concerns	Undetected	No concerns	Some concerns	Some concerns	Some concerns	Low
Pla:TENS	0	No concerns	Undetected	No concerns	Some concerns	No concerns	Some concerns	Moderate
Ros:SMT	0	No concerns	Undetected	No concerns	Some concerns	Some concerns	Some concerns	Low
Ros:TENS	0	No concerns	Undetected	No concerns	Some concerns	No concerns	Some concerns	Moderate
SMT:TENS	0	No concerns	Undetected	No concerns	Some concerns	No concerns	Some concerns	Moderate

3.3 Pain intensity for LBP postpartum

Comparison	Number of studies	Within-study bias	Across-studies bias	Indirectness	Imprecision	Heterogeneity	Incoherence	Confidence rating
Con:Cup	1	No concerns	Undetected	No concerns	Some concerns	Some concerns	No concerns	Moderate
Con:Pla	1	Major concerns	Undetected	No concerns	Some concerns	Some concerns	Some concerns	Very low

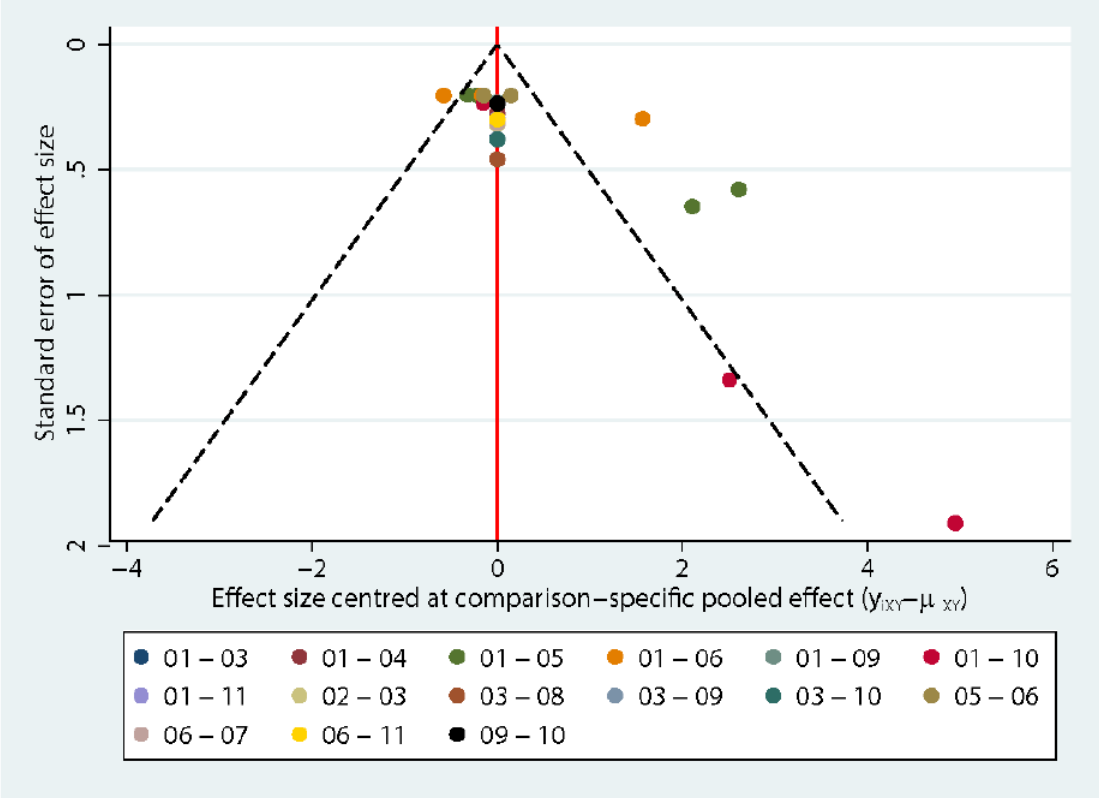
Con:SMT	5	No concerns	Undetected	No concerns	Some concerns	No concerns	No concerns	Moderate
Pla:SMT	1	Major concerns	Undetected	No concerns	Some concerns	Some concerns	Some concerns	Very low
Cup:Pla	0	No concerns	Undetected	No concerns	Major concerns	No concerns	No concerns	Moderate
Cup:SMT	0	No concerns	Undetected	No concerns	Some concerns	Some concerns	No concerns	Moderate

3.4 Physical function for LBP postpartum

Comparison	Number of Studies	Within-Study Bias	Across-Studies Bias	Indirectness	Imprecision	Heterogeneity	Incoherence	Confidence rating
Con:Pla	1	Major concerns	Undetected	No concerns	Some concerns	Some concerns	Some concerns	Very low
Con:SMT	4	No concerns	Undetected	No concerns	No concerns	Major concerns	No concerns	Moderate
Pla:SMT	1	Major concerns	Undetected	No concerns	Some concerns	Some concerns	Some concerns	Very low

GRADE: Grading of Recommendations, Assessment, Development and Evaluations, LBP: low back pain, Exe: exercise, Par: paracetamol, Con: control group, SMT: spinal manipulative treatment, Kin: Kinesio taping, TENS: transcutaneous electrical nerve stimulation, Ros: rose oil, Cup: cupping therapy, PMR: progressive muscle relaxation, Pla: placebo.

Supplementary Figure 2. Comparison-adjusted funnel plot for pain during pregnancy



01: control group, 02: exercise plus paracetamol, 03: paracetamol, 04: progressive muscle relaxation, 05: spinal manipulative treatment, 06: placebo, 07: Kinesio taping, 08: Kinesio taping plus paracetamol, 09: transcutaneous electrical nerve stimulation, 10: exercise, 11: rose oil.

Supplementary Table 4. Results from Direct and Network Evidence [a]

4.1 Physical function during pregnancy

Control group	1.47 (-0.63, 3.58)	2.14 (-0.78, 5.07)	N/A	N/A	**3** **(1.62, 4.38)**	**8** **(6.85, 9.15)**	5.57 (-3.45, 14.59)	**5.12** **(3.91, 6.33)**
1.74 (-0.17, 3.65)	Spinal manipulative treatment	0.33 (-0.04, 0.70)	N/A	N/A	N/A	N/A	N/A	N/A
2.28 **(0.22, 4.35)**	0.54 (-1.74, 2.82)	Placebo	**1.12** **(0.16, 2.08)**	N/A	N/A	N/A	NA	0.66 (-0.55, 1.87)
3.40 (-0.87, 7.67) **Indirect evidence**	1.66 (-2.72, 6.03) **Indirect evidence**	1.12 (-2.61, 4.85)	Kinesio taping	N/A	N/A	N/A	N/A	N/A
7.53 **(2.07, 13.00)** **Indirect evidence**	**5.79** **(0.02, 11.57)** **Indirect evidence**	5.25 (-0.58, 11.08) **Indirect evidence**	4.13 (-2.79, 11.06) **Indirect evidence**	Kinesio taping + Paracetamol	**-3.93** **(-5.38, -2.48)**	N/A	N/A	N/A

3.61 (-0.24, 7.45)	1.86 (-2.41, 6.13) **Indirect evidence**	1.32 (-3.02, 5.67) **Indirect evidence**	0.21 (-5.52, 5.93) **Indirect evidence**	**-3.93** **(-7.82,-0.04)**	Paracetamol	**5** **(3.81, 6.19)**	-1 (-2.41, 0.41)	N/A
8.61 **(4.84, 12.37)**	**6.86** **(2.66, 11.07)** **Indirect evidence**	**6.33** **(2.05, 10.61)** **Indirect evidence**	5.21 (-0.47, 10.89) **Indirect evidence**	1.07 (-4.36,6.51) **Indirect evidence**	**5.00** **(1.20, 8.80)**	TENS	**-6** **(-7.19, -4.81)**	N/A
3.22 (-0.55, 7.00)	1.48 (-2.72, 5.68) **Indirect evidence**	0.94 (-3.33, 5.22) **Indirect evidence**	-0.18 (-5.85, 5.50) **Indirect evidence**	-4.31 (-9.78,1.16) **Indirect evidence**	-0.38 (-4.23, 3.47)	**-5.38** **(-9.16, -1.61)**	Exercise	N/A
4.03 **(0.58, 7.49)**	2.29 (-1.48, 6.06) **Indirect evidence**	1.75 (-1.70, 5.21)	0.63 (-4.46, 5.72) **Indirect evidence**	-3.50 (-9.96,2.96) **Indirect evidence**	0.43 (-4.73, 5.59) **Indirect evidence**	-4.58 (-9.68, 0.53) **Indirect evidence**	0.81 (-4.30, 5.91) **Indirect evidence**	Rose oil

4.2 Treatment withdrawal due to any reason during pregnancy

Control group	0.75 (0.16,3.62)	0.92 (0.18,4.78)	0.72 (0.43,1.21)	0.74 (0.44,1.26)	N/A	0.33 (0.03,3.15)	0.48 (0.04,5.67)	0.32 (0.03,3.18)
0.75 (0.16,3.62)	Progressive muscle relaxation	N/A	N/A	N/A	N/A	N/A	N/A	N/A

1.05 (0.30,3.68)	1.40 (0.19,10.46) **Indirect evidence**	Exercise	N/A	N/A	N/A	1 (0.18,5.6)	1.58 (0.24,10.52)	N/A
0.72 (0.43,1.20)	0.95 (0.18,5.00) **Indirect evidence**	0.68 (0.18,2.64) **Indirect evidence**	Spinal manipulative treatment	1.09 (0.62,1.91)	N/A	N/A	N/A	N/A
0.75 (0.45,1.28)	1.01 (0.19,5.28) **Indirect evidence**	0.72 (0.18,2.79) **Indirect evidence**	1.05 (0.61,1.82)	Placebo	N/A	N/A	N/A	0.65 (0.1,4.11)
0.77 (0.06,9.34) **Indirect evidence**	1.03 (0.05,19.64) **Indirect evidence**	0.74 (0.07,7.65) **Indirect evidence**	1.08 (0.09,13.78) **Indirect evidence**	1.03 (0.08,13.08) **Indirect evidence**	Kinesio taping + Paracetamol	0.97 (0.18,5.16)	N/A	N/A
0.75 (0.12,4.75)	1.00 (0.09,11.31) **Indirect evidence**	0.71 (0.14,3.67)	1.05 (0.15,7.14) **Indirect evidence**	1.00 (0.15,6.77) **Indirect evidence**	0.97 (0.18,5.16)	Paracetamol	1.58 (0.24,10.52)	N/A
1.19 (0.16,8.84)	1.58 (0.12,20.27) **Indirect evidence**	1.13 (0.18,6.96)	1.66 (0.21,13.20) **Indirect evidence**	1.57 (0.20,12.53) **Indirect evidence**	1.53 (0.12,19.19) **Indirect evidence**	1.58 (0.24,10.52)	TENS	N/A
0.42 (0.08,2.28)	0.56 (0.06,5.65) **Indirect evidence**	0.40 (0.05,3.28) **Indirect evidence**	0.59 (0.11,3.29) **Indirect evidence**	0.56 (0.11,2.94)	0.54 (0.03,11.02) **Indirect evidence**	0.56 (0.05,6.84) **Indirect evidence**	0.36 (0.03,4.90) **Indirect evidence**	Rose oil

4.3 Pain intensity postpartum

Spinal manipulative treatment	N/A	**-1.26** **(-1.9, -0.62)**	**-3.06** **(-4.17, -1.95)**
-0.76 (-3.20, 1.68) **Indirect evidence**	Cupping	N/A	**-2.30** **(-2.87, -1.73)**
-1.52 (-3.52, 0.48)	-0.76 (-3.74, 2.23) **Indirect evidence**	Placebo	**-1.27** **(-2.03, -0.51)**
-3.06 **(-4.09, -2.03)**	**-2.30** **(-4.51, -0.09)**	-1.54 (-3.55, 0.46)	Control group

4.4 Physical function postpartum

Spinal manipulative treatment	**-1.34** **(-2.14, -0.54)**	**-2.15** **(-2.84, -1.46)**
-1.03 (-2.26, 0.19)	Placebo	**-1.28** **(-2.08, -0.49)**
-2.20 **(-2.88, -1.51)**	-1.16 (-2.39, 0.06)	Control group

[a] TENS: transcutaneous electrical nerve stimulation

Supplementary Table 5. Rank Results [a]

5.1 Physical function during pregnancy

Rank	Name	SUCRA	Mean Rank (2.5th Percentile, 97.5th Percentile)
1st	Transcutaneous electrical nerve stimulation	94.5	1.4 (1, 3)
2nd	Kinesio taping + Paracetamol	86.8	2.1 (1, 6)
3rd	Rose oil	58.2	4.3 (2, 8)
4th	Paracetamol	50.8	4.9 (3, 9)
5th	Kinesio taping	50.0	5 (2, 9)
6th	Exercise	46.4	5.3 (2, 9)
7th	Placebo	34.3	6.3 (4, 8)
8th	Spinal manipulative treatment	26.3	6.9 (4, 9)
9th	Control group	2.7	8.8 (7, 9)

5.2 Treatment withdrawal due to any reason during pregnancy

Rank	Name	SUCRA	Mean Rank (2.5th Percentile, 97.5th percentile)
1st	Control group	66.7	3.7 (1, 7)
2nd	Transcutaneous electrical nerve stimulation	65.1	3.8 (1, 9)
3rd	Exercise	62.8	4 (1, 8)
4th	Kinesio taping + Paracetamol	48.4	5.1 (1, 9)
5th	Progressive muscle relaxation	47.9	5.2 (1, 9)
6th	Paracetamol	45.8	5.3 (1, 9)
7th	Placebo	45.4	5.4 (2, 9)
8th	Spinal manipulative treatment	41.6	5.7 (2, 9)
9th	Rose oil	26.2	6.9 (1, 9)

5.3 Pain intensity postpartum

Rank	Treatment	SUCRA	Mean Rank (2.5th Percentile, 97.5th Percentile)
1st	Spinal manipulative treatment	88.5	1.3 (1, 2)

2nd	Cupping	64.9	2.1 (1, 3)
3rd	Placebo	43.9	2.7 (1, 4)
4th	Control group	2.7	3.9 (3, 4)

5.4 Physical function postpartum

Rank	Treatment	SUCRA	Mean Rank (2.5th percentile, 97.5th percentile)
1st	Spinal manipulative treatment	97.5	1.0 (1, 2)
2nd	Placebo	51	2.0 (1, 3)
3rd	Control group	1.5	3.0 (2, 3)

[a] SUCRA: surface under the cumulative ranking curve

Supplementary Table 6. Adverse Events of Included Studies

Study ID	Adverse Events
During pregnancy	
Kalinowski 2017	Kinesio taping group: Itching of the area covered by the tape (n=2, 3.8%).
Keskin 2012	Transcutaneous electrical nerve stimulation group: discomfort sense (n=1, 5%). Paracetamol group: gastric intolerance (n=1, 5.3%).
Shirazi 2017	Rose oil group: mild allergic rhinitis (n=1, 2.7%).
Kaplan 2016	Kinesio taping group: allergy (n=2, 6.1%).
Yan et al 2014	Exercise group (n=4, 7.8%): preterm labor (n=2) and uterine contraction too frequencies (n=2). Control group (n=6, 11.8%): preterm labor (n=3), uterine contraction too frequencies (n=2) and bleeding (n=1).
Hensel 2015	Spinal manipulative treatment group (n=11, 8.1%): preterm labor (n=3), preeclampsia (n=3), pregnancy Induced hypertension (n=2), gestational Diabetes (n=1), oligohydramnios (n=1) and eclampsia (n=1). Placebo group (n=19, 14.5%): preterm labor (n=8), preeclampsia (n=3), pregnancy Induced hypertension (n=3), gestational diabetes (n=1), polyhydramnios (n=1), low-lying placenta (n=1), significant 3rd-trimester bleeding (n=1) and premature rupture of membranes (n=1). Control group (n=20, 15%): preterm labor (n=3), preeclampsia (n=4), pregnancy Induced Hypertension (n=4), gestational Diabetes (n=3), oligohydramnios (n=2) and polyhydramnios (n=1), significant 3rd-trimester bleeding (n=1), cardiac dysrhythmia (n=1) and preterm dilation (n=1).
Pekçetin et al 2019	Telephone-supported ergonomic education group: preterm labour (n=1, 1.6%) Face-to-face ergonomic education group: preterm labour (n=2, 3.3%)
Postpartum	
Schwerla 2015	Spinal manipulative treatment group: occasionally, participants reported tired.

Supplementary Table 7. Effect Estimates from Studies which Were not Included in the Network Meta-Analysis [a]

Study ID	Pain	Physical Function
During pregnancy		
Abu 2017	N/A	ODI Exercise + Paracetamol vs Paracetamol: (MD: -2, 95% CI: -4.82– to 0.82)
Akmeses 2014	N/A	SF-36 Progressive muscle relaxation vs Control group: (MD: 65.3, 95% CI: 61.63–68.97)
Mohseni Bandpei et al 2010	N/A	ODI Exercise vs Control group: (MD: -25.74, 95% CI: -45.22 to -6.26)
Kalus 2008	VAS BellyBra vs Tubigrip: (MD: -0.20, 95% CI: -1.10 to 0.70)	Likert scale BellyBra vs Tubigrip: (MD: -0.90, 95% CI: -1.72 to -0.08)
Pekçetin et al 2019	VAS Telephone-supported ergonomic education vs Face-to-face ergonomic education: (MD: -2.19, 95% CI: -3.07 to -1.31)	ODI Telephone-supported ergonomic education vs Face-to-face ergonomic education: (MD: -6.56, 95% CI: -12.65 to -0.47)
Sedaghati 2007	Quebec Exercise vs Control group:	N/A

	(MD: -6.47, 95% CI: -9.22 to -3.72)	
Postpartum		
Mohamed 2018	VAS Kinesio taping + Exercise vs Exercise: (MD: -3.2, 95% CI: -4.04 to -2.36)	BPFS Kinesio taping + Exercise vs Exercise: (MD: 21.5, 95% CI: 18.19 – 24.81)

[a] MD: mean difference, CI: confidence interval, ODI: Oswestry Disability Index, SF-36: 36-Item Short Form Survey, VAS: Visual Analog Scale, Quebec: Quebec questionnaire and BPFS: Back Pain Function Scale

Supplementary Table 8. Sensitivity analyses

8.1 Pain during pregnancy

Interventions	Main Analysis	A	B	C
Progressive muscle relaxation	1st	1st	1st	1st
Kinesio taping	2nd	3rd	3rd	2nd
Kinesio taping + paracetamol	3rd	2nd	2nd	3rd
Transcutaneous electrical nerve stimulation	4th	4th	4th	4th
Exercise + paracetamol	5th	5th	5th	5th
Rose oil	6th	6th	6th	6th
Paracetamol	7th	7th	8th	7th
Exercise	8th	8th	7th	8th
Spinal manipulative treatment	9th	10th	9th	9th
Placebo	10th	9th	10th	10th
Control group	11th	11th	11th	11th

A: exclude studies from grey literature / studies with suspected mixed population (Gundermann 2013 and Roehrich 2014)

B: exclude a study with very high rate of lost to follow-up (Hensel 2015)

C: exclude studies with high risk of bias (Mohseni Bandpei et al 2010)

8.2 Physical function during pregnancy

Interventions	Main Analysis	A	B
Transcutaneous electrical nerve stimulation	1st	1st	1st
Kinesio taping + Paracetamol	2nd	2nd	2nd
Rose oil	3rd	3rd	3rd
Paracetamol	4th	4th	4th
Kinesio taping	5th	6th	6th
Exercise	6th	5th	5th
Placebo	7th	7th	7th
Spinal manipulative treatment	8th	8th	8th
Control group	9th	9th	9th

A: exclude studies from grey literature / studies with suspected mixed population (Gundermann 2013 and Roehrich 2014)

B: exclude a study with very high rate of lost to follow-up (Hensel 2015)

8.3 Pain postpartum

Interventions	Main Analysis	A	B

Spinal manipulative treatment	1st	1st	1st
Cupping	2nd	2nd	2nd
Placebo	3rd	3rd	3rd
Control group	4th	4th	4th

A: exclude studies with suspected mixed population (Schwerla 2015)

B: exclude studies from grey literature (Recknagel 2007 and Belz 2014)

8.4 Physical Function Postpartum

Interventions	Main Analysis	A	B
Spinal manipulative treatment	1st	1st	1st
Placebo	2nd	2nd	2nd
Control group	3rd	3rd	3rd

A: exclude studies with suspected mixed population (Schwerla 2015)

B: exclude studies from grey literature (Recknagel 2007 and Belz 2014)

Supplementary Table 9. Loop and Global Inconsistency [a]

9.1 Pain During Pregnancy

Loop inconsistency			
Loop	IF	95% CI	P value
Con-TENS-Exe	3.46	(1.22, 5.71)	.002
Con-Par-Exe	3.46	(1.07, 5.85)	.005
Con-SMT-Pla	2.02	(0.00, 4.04)	.050
Con-Pla-Ros	1.32	(0.00, 4.76)	.453
Con-Par-TENS	0.00	(0.00, 1.14)	1.000
Par-TENS-Exe	0.00	(0.00, 1.15)	1.000
Global inconsistency			
Chi-square = 73.55, P=0.000			

9.2 Physical function during pregnancy

Loop inconsistency			
Loop	IF	95% CI	P value
Con-Par-Exe	9.50	(0.00, 19.08)	.052
Con-TENS-Exe	9.50	(0.00, 19.02)	.050
Con-Pla-Ros	2.32	(0.00, 10.13)	.561
Con-SMT-Pla	0.56	(0.00, 5.22)	.813
Par-TENS-Exe	0.00	(0.00, 2.20)	1.000
Con-Par-TENS	0.00	(0.00, 2.15)	1.000
Global inconsistency			
Chi-square = 7.18, P=0.0665			

9.3 Pain postpartum

Loop inconsistency			
Loop	IF	95% CI	P value
Con-SMT-Pla	0.663	(0.00, 4.71)	.749
Global inconsistency			
Chi-square = 0.21, P=0.6472			

9.4 Physical function postpartum

Loop inconsistency			
Loop	IF	95% CI	P value

Con-SMT-Pla	0.420	(0.00, 2.84)	.734
Global inconsistency			
Chi-square = 0.19, P=0.6610			

[a] Exe: exercise, Par: paracetamol, Con: control group, SMT: spinal manipulative treatment, Kin: Kinesio taping, TENS: transcutaneous electrical nerve stimulation, Ros: rose oil, Pla: placebo.

Supplementary Table 10. Results of Node-Splitting Method for Pain During Pregnancy [a]

Side	Direct Coef.	Std.Err.	Indirect Coef.	Std.Err.	P value
Par vs Con	-2.00	1.26	-9.12	3.52	0.055
SMT vs Con	-1.55	0.78	-4.21	3.53	0.462
Pla vs Con	-0.97	0.64	-5.58	2.22	0.045
TENS vs Con	-3.00	1.22	-10.19	3.52	0.054
Exe vs Con	-2.73	1.09	4.33	73.12	0.923
Ros vs Con	-3.40	1.51	-0.73	3.29	0.460
Par vs Exe + Par	1.00	1.45	-5.65	93.68	0.943
Kin + Par vs Par	-1.92	1.49	5.75	116.88	0.948
TENS vs Par	-1.00	1.46	-8.06	73.11	0.923
Exe vs Par	1.00	1.26	-6.21	3.53	0.053
Pla vs SMT	-0.01	1.08	1.34	1.90	0.537
Kin vs Pla	-3.71	1.45	2.69	101.55	0.950
Ros vs Pla	-1.12	1.51	-3.79	3.29	0.460
Exe vs TENS	2.00	1.22	-5.22	3.52	0.053

[a] Exe: exercise, Par: paracetamol, Con: control group, SMT: spinal manipulative treatment, Kin: Kinesio taping, TENS: transcutaneous electrical nerve stimulation, Ros: rose oil, Pla: placebo.

Supplementary Table 11. Previous Meta-Analyses [a]

Liddle 2015				
Comparison: during pregnancy - any exercise + usual prenatal care versus prenatal care				
Outcome	No. of Studies	No. of Participants	Statistical Method	Effect Size
Pain intensity	7	645	SMD (IV, Random, 95% CI)	-0.64 (-1.03, -0.25)
Functional disability	2	146	SMD (IV, Random, 95% CI)	-0.56 (-0.89, -0.23)
Franke 2017				
Comparison: during pregnancy – osteopathic manipulative treatment versus control group				
Pain intensity	7	677	MD (IV, Random, 95% CI)	-16.75 (-31.79, -1.72)
Functional disability	7	677	SMD (IV, Random, 95% CI)	-0.50 (-0.93, -0.07)
Comparison: postpartum – osteopathic manipulative treatment versus control group				
Pain intensity	3	173	MD (IV, Random, 95% CI)	-38.00 (-46.75, -29.24)
Functional disability	3	173	SMD (IV, Random, 95% CI)	-2.12 (-3.02, -1.22)
Hall 2016				
Comparison: during pregnancy – manual therapy versus usual care				
Pain intensity	8	865	SMD (IV, Random, 95% CI)	-0.7 (-1.10, -0.30)
Functional disability	5	601	SMD (IV, Random, 95% CI)	-0.62 (-0.93, -0.31)
Number of drop out	4	690	OR (IV, Random, 95% CI)	0.64 (0.20, 2.02)
Comparison: during pregnancy – manual therapy versus relaxation				
Pain intensity	2	82	SMD (IV, Random, 95% CI)	-0.77 (-1.22, -0.32)
Comparison: during pregnancy – manual therapy versus exercise				
Pain intensity	1	57	SMD (IV, Random, 95% CI)	-0.12 (-0.65, 0.42)
Functional disability	1	55	SMD (IV, Random, 95% CI)	-0.21 (-0.77, 0.34)
Number of drop out	1	57	OR (IV, Random, 95% CI)	0.36 (0.10, 1.32)
Comparison: during pregnancy – manual therapy versus sham				
Pain intensity	2	364	SMD (IV, Random, 95% CI)	0.05 (-0.15, 0.26)
Functional disability	2	366	SMD (IV, Random, 95% CI)	-0.08 (-0.40, 0.25)

Number of drop out	2	364	OR (IV, Random, 95% CI)	1.09 (0.62, 1.91)

[a] SMD: standard mean differences, MD: mean differences, IV: inverse variance, OR: odds ratio and CI: confidence interval

CHAPTER THREE

Comparative efficacy and safety of surgical and invasive treatments for adults with degenerative lumbar spinal stenosis: protocol for a network meta-analysis and systematic review

Chapter Three has been published as:

Chen L., H Ferreira P, R Beckenkamp P, L Ferreira M. Comparative efficacy and safety of surgical and invasive treatments for adults with degenerative lumbar spinal stenosis: protocol for a network meta-analysis and systematic review. BMJ Open. 2019 Apr 4;9(4):e024752.

Statement from co-authors confirming authorship contribution of the PhD candidate

The co-authors of the paper "Comparative efficacy and safety of surgical and invasive treatments for adults with degenerative lumbar spinal stenosis: protocol for a network meta-analysis and systematic review" confirm that Lingxiao Chen has made the following contributions:

Conception and design of the research
Analysis and interpretation of the findings
Writing of the manuscript and critical appraisal of the content

In addition to the statements above, in cases where I am not the corresponding author of a published item, permission to include the published material has been granted by the corresponding author.

Lingxiao Chen Date: 21 July 2021

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.
Manuela Loureiro Ferreira Date: 21 July 2021

BMJ Open

Comparative efficacy and safety of surgical and invasive treatments for adults with degenerative lumbar spinal stenosis: protocol for a network meta-analysis and systematic review

Lingxiao Chen,[1] Paulo H Ferreira,[2] Paula R Beckenkamp,[2] Manuela L Ferreira[1]

To cite: Chen L, H Ferreira P, R Beckenkamp P, *et al*. Comparative efficacy and safety of surgical and invasive treatments for adults with degenerative lumbar spinal stenosis: protocol for a network meta-analysis and systematic review. *BMJ Open* 2019;**9**:e024752. doi:10.1136/bmjopen-2018-024752

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2018-024752).

Received 13 June 2018
Revised 14 February 2019
Accepted 15 February 2019

[1]Institute of Bone and Joint Research, Kolling Institute, Sydney Medical School, Faculty of Medicine and Health, University of Sydney, Sydney, New South Wales, Australia
[2]University of Sydney, Faculty of Health Sciences, Discipline of Physiotherapy, Sydney, New South Wales, Australia

Correspondence to
Lingxiao Chen;
lche4036@uni.sydney.edu.au

ABSTRACT

Introduction Surgical and invasive procedures are widely used in adults with degenerative lumbar spinal stenosis when conservative treatments fail. However, little is known about the comparative efficacy and safety of these interventions. To address this, we will perform a network meta-analysis (NMA) and systematic review to compare the efficacy and safety of surgical and invasive procedures for adults with degenerative lumbar spinal stenosis.

Methods and analysis We will include randomised controlled trials assessing surgical and invasive treatments for adults with degenerative lumbar spinal stenosis. We will search AMED, CINAHL, EMBASE, the Cochrane Library and MEDLINE. Only English studies will be included and no restriction will be set for publication status. For efficacy, our primary outcome will be physical function. Secondary outcomes will include pain intensity, health-related quality of life, global impression of recovery, work absenteeism and mobility. For safety, our primary outcome will be all-cause mortality. Secondary outcomes will include adverse events (number of events or number of people with an event) and treatment withdrawal due to adverse effect. Two reviewers will independently select studies, extract data and assess the risk of bias (Revised Cochrane risk-of-bias tool for randomized trials) of included studies. The quality of the evidence will be evaluated through the Grading of Recommendations Assessment, Development and Evaluation framework. Random-effects NMA will be performed to combine all the evidence under the frequentist framework and the ranking results will be presented through the surface under the cumulative ranking curve and mean rank. All analyses will be performed in Stata and R.

Ethics and dissemination No ethical approval is required. The research will be published in a peer-reviewed journal.

PROSPERO registration number CRD42018094180.

Strengths and limitations of this study

- This is the first network meta-analysis to assess the efficacy and safety of surgical and invasive procedures for adults with degenerative lumbar spinal stenosis.
- The main strengths are that only randomised controlled trials will be included for both efficacy outcomes (physical function, pain intensity, health-related quality of life, global impression of recovery, work absenteeism and mobility) and safety outcomes (all-cause mortality, adverse effect and treatment withdrawal due to adverse effect).
- Additional strength is that informative missingness difference of means for continuous outcomes and informative missing ORs for dichotomous outcomes will be used to deal with the missing data.
- The main limitation will be the limited data from lower socioeconomic countries considering the high cost of the surgical and invasive treatments.

INTRODUCTION

Degenerative lumbar spinal canal stenosis is characterised by decreased spinal canal diameter due to structural changes of the spine (eg, facet joints, ligaments) due to ageing. Typically, patients will present with neurogenic claudication, defined as pain, numbness and/or fatigue in the lower limbs that is worsened during walking and standing, and alleviated with forward bending or sitting.[1 2] In the USA, the prevalence of degenerative lumbar spinal stenosis in the general population can be as high as 22.5% for relative stenosis (ie, ≤12 mm canal diameter), and 7.3% for absolute stenosis (ie, ≤10 mm canal diameter).[3] These figures increase drastically with age, reaching 47.2% and 19.4%, respectively, for those 60 years of age or older.[3]

Most guidelines recommend a course of conservative care, including the North American Spine Society guidelines, for patients with degenerative lumbar spinal stenosis.[2] However, when conservative treatments fail, surgical and invasive options are indicated.[2 4 5] Surgical decompression (including laminectomies or laminotomies), with or

without fusion, interspinous process spacer devices, minimally invasive surgical decompression, and corticosteroidal epidural injections are commonly used in the management of spinal stenosis.[6–11] However, the evidence supporting the superiority of one option over the other is still unclear for most.[7 12 13] For instance, past meta-analyses have shown that Superion interspinous spacer is superior to X-STOP interspinous spacer in improving axial pain severity and Zurich Claudication Questionnaire (ZCQ) patient satisfaction score; whereas the addition of spinal fusion to surgical decompression does not add any benefit to surgical decompression alone.[14 15] Moreover, existing meta-analyses use pairwise analytical approaches, and therefore can only provide results for the comparison of two interventions at any one time.[4 11 14–28] A network meta-analysis (NMA) is the best design and analytical approach to compare and rank multiple interventions simultaneously, based on their relative estimate effects in each outcome.[29] NMA has been used in similar fields, including sciatica, lumbar disc herniation and osteoarthritis, but, to date, no NMA has been conducted to establish the comparative effectiveness and safety of invasive approaches for degenerative lumbar spinal canal stenosis.[30–32] As such, our aim is to perform an NMA and systematic review to assess the efficacy and safety of surgical and invasive procedures for adults with degenerative lumbar spinal stenosis.

METHODS AND ANALYSIS

Criteria for considering studies for this review

The protocol was written in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols.[33] Any changes made to this protocol will be updated in the PROSPERO registration.

Types of participants

We will include studies that recruited participants who are 40 years of age or older, with a diagnosis of degenerative lumbar spinal stenosis. We will exclude studies on patients with malignancy, trauma, vertebral fracture, infection and inflammatory disorders. For studies including degenerative lumbar spinal stenosis and associated spondylolisthesis, only those of participants with Meyerding grade I spondylolisthesis will be included. Studies including mixed populations will only be included if the data for patients with degenerative lumbar spinal stenosis can be extracted separately or if at least 80% of the patients are diagnosed with degenerative lumbar spinal stenosis.

Types of interventions

Studies comparing any surgical or invasive intervention for adults with degenerative lumbar spinal stenosis will be included. For example, surgical decompression, including laminectomies or laminotomies, with or without fusion, interspinous process spacer devices, minimally invasive surgical decompression and corticosteroidal epidural injections. The comparison group could be no treatment, usual care, sham operation, another active option or a combination of approaches. The interventions in comparison groups will be treated as different nodes. However, if we have insufficient studies to connect different interventions, we will combine no treatment and usual care into one node to make full use of the data.

Outcome measures

The outcome data will be grouped into short-term (≤6 months), mid-term (6–12 months) and long-term (≥12 months) follow-up assessment.[34] We will perform NMA in the three time points separately. For studies which report outcomes in multiple time points, data closest to the 6 and 12 months follow-up time will be included in the main analyses. For different time points in long-term follow-up assessment (eg, 1 year, 2 years, 5 years), subgroup analyses will be performed.

Primary outcomes

1. Physical function, commonly measured by Oswestry Disability Index (ODI), Roland Morris Disability Questionnaire (RMDQ), Patient-Specific Function Scale and Core Outcome Measures Index (COMI).[34] Other rating scales will be included if they have been proposed in peer-reviewed journals. If the study provides more than one instruments, ODI will be used as the first choice, RMDQ as the second choice and COMI as the third choice.[34]
2. All-cause mortality measured by the percentage of patients who died following randomisation.

Secondary outcomes

1. Pain intensity, commonly measured by Numeric Rating Scale (NRS) and the Visual Analogue Scale (VAS).[35 36] Other rating scales will also be included if they have been proposed in peer-reviewed journals. Pain intensity will be categorised and analysed according to the following three groups: back pain, leg pain and overall pain. If the study provides more than one instruments, VAS will be used as the first choice and NRS as the second choice.[34]
2. Health-related quality of life, commonly measured by 36-Item Short Form Survey (SF-36), EuroQol five-dimension (EQ-5D), Nottingham health profile (NHP) and SF-12.[34] SF-36, NHP and SF-12, could be mapped into EQ-5D.[37] As above, other tools will also be included if they have been proposed in peer-reviewed journals. If the study provides more than one instruments, EQ-5D will be used as the first choice, following by SF-36, SF-12 and NHP.[34]
3. Global impression of recovery measured by the percentage of the patients satisfied with their recovery.
4. Work absenteeism measured by the number of days of sick leave.
5. Mobility measured by walking distance.
6. Adverse event measured by the number of participants with an adverse event, or number of adverse events per group. Adverse events could include nerve injury, du-

ral tear, vascular injury, deep infection and pulmonary embolus.
7. Treatment withdrawal due to adverse effect measured by the percentage of patients who drop out due to adverse effect.

Types of studies

Only randomised controlled trials, which include parallel, cross-over and cluster trials, will be included. For cross-over studies, only data before wash-out period will be used. For cluster randomised trials, we will extract data which are adjusted for clustering. If these data are unavailable, we will extract original data and adjust them for clustering.[38 39] To decrease bias, we excluded studies with a high risk of bias in the domain risk of bias arising from the randomisation process.[40]

Search strategy

Electronic searches

The following databases will be searched for published studies: AMED, CINAHL, EMBASE, the Cochrane Library and MEDLINE (including MEDLINE Epub Ahead of Print, In-Process and Other Non-Indexed Citations, MEDLINE Daily and MEDLINE). Unpublished and ongoing studies will be searched from WHO International Clinical Trials Registry Platform (http://www.who.int/ictrp/en/) and the US National Institutes of Health (https://clinicaltrials.gov/). Only English studies will be included and no restriction will be set for publication status. The search strategy for MEDLINE is provided as online supplementary material.

Reference lists and other sources

Reference lists of all included studies, relevant systematic reviews and meta-analyses, and guidelines will be screened for eligible additional studies to be included.

Identification and selection of studies

Two reviewers will independently screen titles and abstracts of the articles from the search. Before formal screening of titles, we will perform an intratester agreement test (kappa test) by randomly selecting 50 citations (through random number table) to be reviewed by two independent reviewers.[38] An agreement of 80% or more will be considered acceptable. If we do not achieve the percentage of the agreement, we will randomly select another 50 citations subsequently until 80% of agreement is reached. Any disagreement will be solved by discussion and if necessary, a third reviewer will arbitrate the decision. When studies fail to provide the necessary data, the authors will be contacted and further information requested.

Data extraction

Two reviewers will independently extract data from the included studies using a standardised data extraction form. Similarly, a pilot test will be performed before the formal extraction. We will randomly select five articles using a random number table to confirm we have enough inter-rater agreement (at least 80%). Any disagreement will be solved by discussion. Otherwise, a third reviewer will make a decision. The following data will be extracted from each included study based on recommendations from previous studies.[34 41]

1. Study characteristics, such as year of study publication, first author, journal, sample size, study funding and location.
2. Patient characteristics, such as age, gender, including and excluding criteria, diagnostic criteria, type of lumbar spinal stenosis, comorbidities, duration of symptoms and previous treatment.
3. Intervention characteristics.
4. Primary and secondary outcomes.

Measurement of treatment effect

Relative treatment effects

1. Continuous outcomes: If the studies use the same rating scale, we will use mean difference (MD) with its 95% CI. If different rating scales are used, standardised MD with its 95% CI will be used.
2. Dichotomous outcomes: OR with its 95% CI will be used.
3. For all-cause mortality, the number needed to harm will be calculated.[38]

Relative treatment ranking

The surface under the cumulative ranking curve and mean ranks will be used to rank each intervention for each outcome.[42] Rank-heat plot will be used to show the ranking results of each outcome for each intervention.[43]

Dealing with missing outcome data and missing statistics

For continuous outcomes, if the study only reports SE, p value or CI, we will convert them into SD.[38] If the study reports median and IQR, we will calculate SD by dividing the IQR by 1.35 and considering the median equivalent to the mean.[38] If relevant information is provided in figures, we will extract the data from the graphs. If data cannot be obtained, we will contact the authors. If we do not obtain relevant data, informative missingness difference of means (IMDoM) will be used as one kind of sensitivity analysis to explore the uncertainty of our results under the missing at random assumption.[44]

For dichotomous outcomes, first, we will try to contact the authors to obtain data. In the absence of a response or of relevant data, informative missing ORs (IMORs) for dichotomous outcomes will be used to explore the uncertainty of our results under the missing at random assumption.[44]

Risk of bias assessment

Two reviewers will independently assess the risk of bias of the included studies. Any disagreement will be solved by discussion. Otherwise, a third reviewer will make a decision. We will contact the authors to obtain further information if the third reviewer thinks it is necessary.

Revised Cochrane risk-of-bias tool for randomized trials (RoB 2) will be used to evaluate the risk of bias

of included randomised parallel-group trials.[40] The tool is composed of five domains: (1) bias arising from the randomisation process; (2) bias due to deviations from intended interventions; (3) bias due to missing outcome data; (4) bias due to missing outcome data and (5) bias in selection of the reported result. Each domain includes several signalling questions which elicit information relevant to an assessment of risk of bias. The answer option for each signalling question is: yes, probably yes, probably no, no and no information. Based on the answers of all signalling questions in one domain, we will rate the domain as low risk of bias, some concerns or high risk of bias. Finally, we will get the overall risk-of-bias judgement as low risk of bias, some concerns or high risk of bias considering the risk-of-bias judgement in five domains.

For cluster-randomised trials, one more domain should be considered: bias arising from identification or recruitment of individual participants within clusters. For cross-over trials, analysis issues in cross-over trials should be additionally considered.

Data analysis

The characteristics of study, patient and intervention will be summarised descriptively. We will make a narrative review for some comparisons if insufficient data are provided. Network plot will be drawn to descript the available interventions. The size of the node reflects the number of patients in each intervention. The breadth of the edge shows the number of comparisons. For efficacy and safety outcomes, pairwise and NMA will be performed.

Pairwise meta-analyses

We will perform traditional pairwise meta-analyses through random-effect model with DerSimonian and Laird inverse-variance method for every direct comparison.[38] In some subgroups, we will also perform pairwise meta-analyses if NMAs could not be performed. The heterogeneity will be assessed by I^2 and T^2.[38]

Assessment of the transitivity assumption

The potential baseline effect modifiers (age, gender, education level, baseline physical function, smoking habit, body mass index (BMI), comorbidities and previous treatment) will be assessed to confirm they are similar among different comparisons before we perform NMAs.[34] If any difference is found, we will perform meta-regression to explore the influence on the results.

Network meta-analyses

Random-effect NMAs under the frequentist framework will be performed to combine both direct and indirect comparisons.[45] The heterogeneity parameter is assumed the same for each intervention.[45] Prediction interval plot will be drawn to reflect the uncertainly of the results in a future study.[46 47]

Assessment of inconsistency

Bucher method as a local method and design-by-treatment interaction model as a global method will be used.[48 49] If any inconsistency is found, node-splitting method will be used to explore the original of the inconsistency.[45]

Exploring sources of heterogeneity or inconsistency with subgroup analyses and meta-regression

For two primary outcomes (physical function and all-cause mortality), subgroup analyses and meta-regressions will be performed under the three-time categories (short term, mid term and long term) except for the analysis on duration of follow-up for long-term assessment. Subgroup analyses will be performed as follows: (1) Single-level spinal stenosis versus multiple levels, the hypothesis is that patients with multiple levels spinal stenosis might have poorer physical function and higher all-cause mortality than patients with single level; (2) Duration of follow-up for long-term assessment (eg, 1 year, 2 years and 5 years), the hypothesis is that patients who received injection therapies might have poorer physical function and higher all-cause mortality in longer duration of follow than patients who received surgical therapies; (3) Patients with versus patients without degenerative spondylolisthesis, the hypothesis is that patients with degenerative spondylolisthesis might have poorer physical function and higher all-cause mortality than patients without; (4) Type of disease: central, foraminal or lateral, the hypothesis is that patients with central lumbar spinal stenosis might have poorer physical function and higher all-cause mortality than patients with foraminal or lateral. Meta-regression will be performed as follows: (1) age; (2) percentage of the male; (3) sample size; (4) baseline physical function; (5) percentage of the smoker and (6) BMI.

Sensitivity analyses

For two primary outcomes (physical function and all-cause mortality), sensitivity analyses will be performed as follows: (1) only studies with low risk of bias; (2) studies with imputed data through either IMDoM or IMOR; (3) studies without a non-active comparison group; (4) studies without receiving commercial funding and (5) studies without unpublished data.

Publication bias

Comparison-adjusted funnel plot will be used to test the publication bias if the number of included studies is larger than 10.[42] As described above, meta-regression procedures using sample size and effect estimates will be performed to detect the small-study effect.[50]

Grading the evidence

The Grading of Recommendations, Assessment, Development and Evaluations framework will be used to evaluate the quality of evidence.[51] The tool includes five domains, which are study limitations, indirectness, inconsistency, imprecision and publication bias.

Statistical software

All analyses (pairwise meta-analysis will be only performed in Stata and NMA will be performed in both Stata and R) will be performed in Stata (StataCorp. 2017. Stata Statistical Software: Version 15.1) and R (V.3.4.3. R Core Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria).

Patient and public involvement

Patients will not be involved.

ETHICS AND DISSEMINATION

We will publish the research in a peer-reviewed journal after completing it.

Contributors All authors conceived the study. LC drafted the manuscript. LC and PB participated in the search strategy development. PF, PB and MF assisted in protocol design and revision. All authors read and approved the final manuscript as submitted.

Funding MLF is supported by the National Health and Medical Research Council of Australia Career Development Fellowship. Other authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval This research does not require ethics approval because it uses data from literatures.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. Hartvigsen J, Hancock MJ, Kongsted A, *et al*. What low back pain is and why we need to pay attention. *Lancet* 2018;391:2356–67.
2. Kreiner DS, Shaffer WO, Baisden JL, *et al*. An evidence-based clinical guideline for the diagnosis and treatment of degenerative lumbar spinal stenosis (update). *Spine J* 2013;13:734–43.
3. Kalichman L, Cole R, Kim DH, *et al*. Spinal stenosis prevalence and association with symptoms: the Framingham Study. *Spine J* 2009;9:545–50.
4. Machado GC, Ferreira PH, Yoo RI, *et al*. Surgical options for lumbar spinal stenosis. *Cochrane Database Syst Rev* 2016;11:Cd012421.
5. May S, Comer C. Is surgery more effective than non-surgical treatment for spinal stenosis, and which non-surgical treatment is more effective? A systematic review. *Physiotherapy* 2013;99:12–20.
6. Benyamin RM, Staats PS, MiDAS Encore I. MILD® Is an Effective Treatment for Lumbar Spinal Stenosis with Neurogenic Claudication: MiDAS ENCORE Randomized Controlled Trial. *Pain Physician* 2016;19:229–42.
7. Deyo RA, Martin BI, Ching A, *et al*. Interspinous spacers compared with decompression or fusion for lumbar stenosis: complications and repeat operations in the Medicare population. *Spine* 2013;38:865–72.
8. Deyo RA, Mirza SK, Martin BI, *et al*. Trends, major medical complications, and charges associated with surgery for lumbar spinal stenosis in older adults. *JAMA* 2010;303:1259–65.
9. Elsheikh NA, Amr YM. Effect of adding calcitonin to translaminar epidural steroid in degenerative lumbar spinal canal stenosis. *Pain Physician* 2016;19:139–46.
10. Jansson KA, Blomqvist P, Granath F, *et al*. Spinal stenosis surgery in Sweden 1987-1999. *Eur Spine J* 2003;12:535–41.
11. Manchikanti L, Knezevic NN, Boswell MV, *et al*. Epidural injections for lumbar radiculopathy and spinal stenosis: a comparative systematic review and meta-analysis. *Pain Physician* 2016;19:E365–410.
12. Försth P, Ólafsson G, Carlsson T, *et al*. A randomized, controlled trial of fusion surgery for lumbar spinal stenosis. *N Engl J Med* 2016;374:1413–23.
13. Ghogawala Z, Dziura J, Butler WE, *et al*. Laminectomy plus Fusion versus Laminectomy Alone for Lumbar Spondylolisthesis. *N Engl J Med* 2016;374:1424–34.
14. Shen J, Xu S, Xu S, *et al*. Fusion or Not for Degenerative Lumbar Spinal Stenosis: A Meta-Analysis and Systematic Review. *Pain Physician* 2018;21:1–8.
15. Zhao H, Duan LJ, Gao YS, *et al*. Comparison of two FDA-approved interspinous spacers for treatment of lumbar spinal stenosis: Superion versus X-STOP-a meta-analysis from five randomized controlled trial studies. *J Orthop Surg Res* 2018;13:42.
16. Zhao XW, Ma JX, Ma XL, *et al*. Interspinous process devices(IPD) alone versus decompression surgery for lumbar spinal stenosis(LSS): A systematic review and meta-analysis of randomized controlled trials. *Int J Surg* 2017;39:57–64.
17. Yavin D, Casha S, Wiebe S, *et al*. Lumbar fusion for degenerative disease: a systematic review and meta-analysis. *Neurosurgery* 2017;80:701–15.
18. Li AM, Li X, Yang Z. Decompression and coflex interlaminar stabilisation compared with conventional surgical procedures for lumbar spinal stenosis: A systematic review and meta-analysis. *Int J Surg* 2017;40:60–7.
19. Lee CH, Chung CK, Sohn MJ, *et al*. Short limited fusion versus long fusion with deformity correction for spinal stenosis with balanced De Novo degenerative lumbar scoliosis: a meta-analysis of direct comparative studies. *Spine* 2017;42:E1126–32.
20. Chang W, Yuwen P, Zhu Y, *et al*. Effectiveness of decompression alone versus decompression plus fusion for lumbar spinal stenosis: a systematic review and meta-analysis. *Arch Orthop Trauma Surg* 2017;137:637–50.
21. Phan K, Rao PJ, Ball JR, *et al*. Interspinous process spacers versus traditional decompression for lumbar spinal stenosis: systematic review and meta-analysis. *J Spine Surg* 2016;2:31–40.
22. Phan K, Mobbs RJ. Minimally invasive versus open laminectomy for lumbar stenosis: a systematic review and meta-analysis. *Spine* 2016;41:E91–100.
23. Overdevest GM, Jacobs W, Vleggeert-Lankamp C, *et al*. Effectiveness of posterior decompression techniques compared with conventional laminectomy for lumbar stenosis. *The Cochrane database of systematic reviews* 2015;3:Cd010036.
24. Machado GC, Ferreira PH, Harris IA, *et al*. Effectiveness of surgery for lumbar spinal stenosis: a systematic review and meta-analysis. *PLoS One* 2015;10:e0122800.
25. Liu K, Liu P, Liu R, *et al*. Steroid for epidural injection in spinal stenosis: a systematic review and meta-analysis. *Drug Des Devel Ther* 2015;9:707–16.
26. Liang L, Jiang WM, Li XF, *et al*. Effect of fusion following decompression for lumbar spinal stenosis: a meta-analysis and systematic review. *Int J Clin Exp Med* 2015;8:14615–24.
27. Lauryssen C, Jackson RJ, Baron JM, *et al*. Stand-alone interspinous spacer versus decompressive laminectomy for treatment of lumbar spinal stenosis. *Expert Rev Med Devices* 2015;12:763–9.
28. Hong P, Liu Y, Li H. Comparison of the efficacy and safety between interspinous process distraction device and open decompression surgery in treating lumbar spinal stenosis: a meta analysis. *J Invest Surg* 2015;28:40–9.
29. Dias S, Welton NJ, Sutton AJ, *et al*. Evidence synthesis for decision making 1: introduction. *Med Decis Making* 2013;33:597–606.
30. da Costa BR, Reichenbach S, Keller N, *et al*. Effectiveness of non-steroidal anti-inflammatory drugs for the treatment of pain in knee and hip osteoarthritis: a network meta-analysis. *Lancet* 2017;390:e21–33.
31. Feng F, Xu Q, Yan F, *et al*. Comparison of 7 surgical interventions for lumbar disc herniation: a network meta-analysis. *Pain Physician* 2017;20:E863–71.
32. Guo JR, Jin XJ, Shen HC, *et al*. A comparison of the efficacy and tolerability of the treatments for sciatica: a network meta-analysis. *Ann Pharmacother* 2017;51:1041–52.
33. Moher D, Shamseer L, Clarke M, *et al*. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev* 2015;4:1.
34. Clement RC, Welander A, Stowell C, *et al*. A proposed set of metrics for standardized outcome reporting in the management of low back pain. *Acta Orthop* 2015;86:523–33.

35. Chapman JR, Norvell DC, Hermsmeyer JT, *et al*. Evaluating common outcomes for measuring treatment success for chronic low back pain. *Spine* 2011;36(21 Suppl):S54–68.
36. Jensen MP, Mardekian J, Lakshminarayanan M, *et al*. Validity of 24-h recall ratings of pain severity: biasing effects of "Peak" and "End" pain. *Pain* 2008;137:422–7.
37. Dakin H, Abel L, Burns R, *et al*. Review and critical appraisal of studies mapping from quality of life or clinical measures to EQ-5D: an online database and application of the MAPS statement. *Health Qual Life Outcomes* 2018;16:31.
38. The Cochrane collaboration. *Cochrane handbook for systematic reviews of interventions*: The Cochrane collaboration, 2011. Available from. www.handbook.cochrane.org.
39. Caddy C, Amit BH, McCloud TL, *et al*. Ketamine and other glutamate receptor modulators for depression in adults. *Cochrane Database Syst Rev* 2015:Cd011612.
40. Chandler J, McKenzie J, Boutron I, Welch V, *et al*. eds. *Cochrane methods*: Cochrane Database of Systematic Reviews, 2016.
41. Furlan AD, Malmivaara A, Chou R, *et al*. 2015 Updated Method Guideline for Systematic Reviews in the Cochrane Back and Neck Group. *Spine* 2015;40:1660–73.
42. Salanti G, Ades AE, Ioannidis JP. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. *J Clin Epidemiol* 2011;64:163–71.
43. Veroniki AA, Straus SE, Fyraridis A, *et al*. The rank-heat plot is a novel way to present the results from a network meta-analysis including multiple outcomes. *J Clin Epidemiol* 2016;76:193–9.
44. Mavridis D, White IR, Higgins JP, *et al*. Allowing for uncertainty due to missing continuous outcome data in pairwise and network meta-analysis. *Stat Med* 2015;34:721–41.
45. White IR. Network meta-analysis. *Stata J* 2015;15:951–85.
46. Riley RD, Higgins JP, Deeks JJ. Interpretation of random effects meta-analyses. *BMJ* 2011;342:d549.
47. Higgins JP, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *J R Stat Soc Ser A Stat Soc* 2009;172:137–59.
48. Bucher HC, Guyatt GH, Griffith LE, *et al*. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol* 1997;50:683–91.
49. Higgins JP, Jackson D, Barrett JK, *et al*. Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. *Res Synth Methods* 2012;3:98–110.
50. Chaimani A, Salanti G. Using network meta-analysis to evaluate the existence of small-study effects in a network of interventions. *Res Synth Methods* 2012;3:161–76.
51. Salanti G, Del Giovane C, Chaimani A, *et al*. Evaluating the quality of evidence from a network meta-analysis. *PLoS One* 2014;9:e99682.

CHAPTER FOUR

Effectiveness and Safety of Surgical, Invasive Treatments and Conservative Care for Degenerative Lumbar Spinal Stenosis: A Systematic Review and Network Meta-analysis

Chapter Four has been submitted for publication as:

Chen L., Ferreira P.H., Stanford R., Beckenkamp P.R., Bayartai M., Anderson D.B., Arden N.K., Ferreira M.L**.** Effectiveness and Safety of Surgical, Invasive Treatments and Conservative Care for Degenerative Lumbar Spinal Stenosis: A Systematic Review and Network Meta-analysis. *Under review*

Statement from co-authors confirming authorship contribution of the PhD candidate

The co-authors of the paper "Effectiveness and Safety of Surgical, Invasive Treatments and Conservative Care for Degenerative Lumbar Spinal Stenosis: A Systematic Review and Network Meta-analysis" confirm that Lingxiao Chen has made the following contributions:

Conception and design of the research
Analysis and interpretation of the findings
Writing of the manuscript and critical appraisal of the content

In addition to the statements above, in cases where I am not the corresponding author of a published item, permission to include the published material has been granted by the corresponding author.

Lingxiao Chen Date: 21 July 2021

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.
Manuela Loureiro Ferreira Date: 21 July 2021

Effectiveness and Safety of Surgical, Invasive Treatments and Conservative Care for Degenerative Lumbar Spinal Stenosis: A Systematic Review and Network Meta-analysis

Lingxiao Chen MMed[1], Paulo H Ferreira PhD[2], Ralph Stanford MD[3], Paula R Beckenkamp PhD[2], Munkh-erdene Bayartai MD[2], David B Anderson Msc[1], Nigel K Arden MD[4], Manuela L Ferreira PhD[1]

[1]Institute of Bone and Joint Research, The Kolling Institute, Northern Clinical School, Faculty of Medicine and Health, University of Sydney, Sydney, New South Wales, Australia

[2]University of Sydney, School of Health Sciences, Faculty of Medicine and Health, Sydney, New South Wales, Australia

[3]Prince of Wales Hospital, Sydney, New South Wales, Australia

[4]Centre for Sport, Exercise and Osteoarthritis Research Versus Arthritis, University of Oxford, Oxford, UK.

Correspondence to: Lingxiao Chen

lche4036@uni.sydney.edu.au

(or @LingxiaoChen2 on Twitter; Tel:610466965326)

ABSTRACT

Importance There are many surgical and invasive interventions available to treat degenerative lumbar spinal stenosis, but we are still not clear whether surgical or invasive interventions are better than conservative care (e.g., exercise). Degenerative lumbar spinal stenosis is heterogeneous which often incurs debates about the choice of treatments between physicians and surgeons.

Objective To compare the effectiveness and safety of surgical interventions or invasive treatments to conservative care for degenerative lumbar spinal stenosis.

Data Sources AMED, CINAHL, EMBASE, the Cochrane Library and MEDLINE (inception to September 2020), trial registers, and reference lists of included studies.

Study Selection Randomised controlled trials evaluating surgical or invasive interventions (e.g., epidural injections) for the treatment of degenerative lumbar spinal stenosis in adults aged 40 years or older.

Data Extraction and Synthesis Two authors independently extracted the data. Frequentist network and pairwise meta-analyses were performed. All data analyses were performed in either short-term (⩽6 months) or long-term (⩾12 months) separately if applicable.

Main Outcomes and Measures The primary effectiveness outcome was physical function, measured by the Oswestry Disability Index, Roland Morris Disability Questionnaire, Japanese Orthopaedic Association score, functional rate index or neurogenic claudication outcome score. The primary adverse outcome was all-cause mortality.

Results A total of 49 trials (mean age ranged from 52 to 76) with 5323 patients and 16 interventions were included. For short-term (26 trials with 3247 patients and 13 interventions) and long-term (27 trials with 3342 patients and 9 interventions) physical function, no statistical difference was observed when surgical or invasive treatments were compared with conservative care. For all-cause mortality (10 trials with 1573 patients and 6 interventions), there was no statistical difference between any surgical treatment and conservative care.

Conclusions and Relevance There was no evidence to support that surgery or invasive procedures are more effective or safer than conservative care in treating degenerative lumbar spinal stenosis.

This information may inform discussion with patients about the treatment options in degenerative lumbar spinal stenosis.

Introduction

Degenerative lumbar spinal stenosis (LSS) is a narrowing of the spinal canal diameter due to degenerative changes[1]. Patients with central LSS typically present with neurogenic claudication, which is defined as pain, numbness and/or fatigue in the lower limbs[2] that is worsened during walking and standing (i.e., lumbar extension), and alleviated with forward bending or sitting (i.e., lumbar flexion)[1]. Results from the Framingham study, based on radiology, indicated that the prevalence of absolute LSS (i.e., ⩽10 mm canal diameter) and relative LSS (i.e., ⩽12 mm canal diameter) in the general population is 7.3% and 22.5%, respectively[3]. For those aged 60 years and older, the prevalence of absolute LSS and relative LSS is higher and can reach 19.4% and 47.2%, respectively[3].

Besides conservative care (i.e., exercise), therapeutic procedures such as surgical interventions, including surgical decompression with or without fusion, interspinous process spacer devices, minimally invasive surgical decompression and epidural injections are commonly used to treat degenerative LSS[4-6]. To date, there is limited evidence to support one treatment type over another, with a Cochrane review concluding no superiority of surgery over non-surgical care for self-reported disability at 12 months (mean difference (MD): -6.18, 95% confidence interval (CI) -15.3 to 2.66) and a statistically significant but clinically unimportant difference at 24 months (MD: -4.43, 95% CI -7.91 to -0.96). A limitation of the pairwise Cochrane review finding was the small number of studies available, with only five randomized controlled trials (RCTs) included[7]. An additional Cochrane review compared three different posterior decompression techniques to conventional laminectomy and found no significant difference between treatments, but was also limited by the small number of studies available for inclusion[8]. Moreover, previous studies using pair-wise analytic approaches, have failed to provide the overall comparative effectiveness and safety among all surgical and non-surgical procedures for lumbar spinal stenosis. A network meta-analysis approach can, however, simultaneously compare the effectiveness of all included interventions providing comparisons not possible in previous meta-analyses. Furthermore, compared to pair-wise approaches, the network approach enables the inclusion of additional trials in the analyses, increasing the precision of the final results [9]. We have therefore conducted a

systematic review with a network meta-analysis, to assess the effectiveness and safety of surgical or invasive treatments and conservative care for people with degenerative LSS.

Methods

The reporting of the study followed the PRISMA Extension Statement for Reporting of Systematic Reviews Incorporating Network Meta-analyses of Health Care Interventions and was registered in PROSPERO (CRD42018094180)[10]. The protocol for this paper was published elsewhere[11].

Eligibility criteria

Population

To be included, trials needed to recruit participants who were aged 40 years or older with a diagnosis of degenerative LSS (either central, lateral and foraminal). Details could be found in Appendix S1.

Intervention, comparison, and study design

Trials were included if they randomised participants to a surgical (e.g., decompression) or invasive (i.e., epidural injection) intervention compared to any other treatment (e.g., conservative care, another active option), and were published in English. All types of conservative care (included one or multiple components from the following options: exercise, education, and nonsteroidal anti-inflammatory drugs if needed) were treated as one node and two sensitivity analyses were performed to verify the robustness of this strategy (Appendix S2).

Outcome Measures

The primary outcomes were:

Physical function as a primary effectiveness outcome, measured by the Oswestry Disability Index, Roland Morris Disability Questionnaire, Japanese Orthopaedic Association score, functional rate index or neurogenic claudication outcome score;

All-cause mortality as a primary adverse outcome.
The secondary outcomes were:

Back pain;
Leg pain;
Overall pain (not specific to a body part);
Global impression of recovery, defined as satisfaction rate;
Mobility, measured by walking distance without pain, maximum distance of 900 m, the distance in metres that the patient could walk on even ground without a break or 6-minute walk test;
Adverse effect (other than mortality) due to any reason;
Intervention related adverse effect;
Reoperation rate;
Treatment withdrawal (due to any reason);
Treatment withdrawal (due to adverse effects);
Health-related quality of life;
Work absenteeism
Treatment withdrawal (due to adverse effects), health-related quality of life and work absenteeism were not reported given the insufficient data available to perform meta-analyses. Intervention related adverse events were added as secondary outcome measure to provide further information on the type and relatedness of the adverse event. Moreover, data on reoperation rate were extracted and reported as a secondary outcome (detailed definition for these added outcomes were provided in Appendix S3).

Data Sources and Searches

We searched AMED, CINAHL (via Ebsco), EMBASE (via OvidSP), the Cochrane Library and MEDLINE (via OvidSP) for published studies from inception to March 20th, 2019, with updated

searches conducted on the 5th of March, 2020 and 5th of September, 2020. Detailed search strategies, based on the key inputs from a trained librarian, could be found in published protocol[11].

Study Selection, Data Extraction and Risk of Bias Assessment

Two reviewers (LC and MB) independently selected eligible studies through screening their titles, abstracts and full-text articles. Any disagreement was judged by the senior author (MF) and clinical expert (RS). The senior author (MF) made the final decision if any disagreement still existed. We extracted data on study characteristics (e.g., publication year, geographical region, study duration and funding source), patient characteristics (e.g., mean age, sex ratio, stenosis level and stenosis type), and intervention types. Two reviewers (LC and MB) independently assessed the risk of bias of included studies through revised Cochrane risk-of-bias tool for randomised trials (RoB 2)[12].

Data Synthesis and Analysis

One reviewer (LC) performed all data analysis. A frequentist framework was used to conduct pairwise and network meta-analyses in Stata, version 15[13-15]. Random effects model was used to account for potential methodological and clinical heterogeneity. The heterogeneity parameter was assumed the same to increase power in the heterogeneity estimation. Continuous outcomes using the same rating scale (back pain, leg pain and overall pain) were reported as mean differences (MD) with its 95% confidence interval (CI). Continuous outcomes using different rating scale (physical function and mobility) were reported as standardised mean difference (SMD) with 95% CI. Dichotomous outcomes (all-cause mortality, global impression of recovery, adverse event due to any reason, intervention related adverse event, reoperation rate and treatment withdrawal due to any reason) were reported as odds ratio (OR) with its 95% CI.

The magnitude of effects (small, moderate and large) for pain intensity and physical function was defined from the recommendations from American College of Physicians Clinical Practice Guideline (Appendix S4) [16,17]. In this study, the term clinical importance was defined by at least a moderate effect (i.e., 10 in 0-100 pain score system and 0.5 in SMD for physical function) for the

lower confidence limit of a positive value and the upper confidence limit of a negative value[17]. As our protocol indicated, all data analyses were performed in either short-term (≤6 months) or long-term (≥12 months) separately considering we have insufficient data in middle-term (6-12 months).

To examine the transitivity assumption, we visually inspected the central tendencies of the study and patient characteristics[13]. Design-by-treatment interaction model and node-splitting method were used to assess the inconsistency[15]. Design-by-treatment interaction model could assess whether the particular choice of treatments in a study is associated with different effect sizes for particular contrasts when multi-arm trials exist[18]. Node-splitting method could split one treatment contrast into the direct parameter and the indirect parameter, and then test whether the difference exists between two parameters[19]. To rank all available interventions in each outcome, we used the Surface Under the Cumulative Ranking curves (SUCRA) and mean rank[20]. For two primary outcomes, we used comparison-adjusted funnel plot to assess the potential publication bias[21].

We performed network meta-analyses only if the network plot included conservative care as a comparison to provide more clinically relevant results. The only exception was with reoperation rates which was only analysed across surgical comparisons. To examine the robustness of the results from two primary outcomes, extensive sensitivity analyses were performed (Details could be in Appendix S5 and S6). All analyses were performed in STATA 15.1.

Quality of the evidence appraisal

Two reviewers independently used Grading of Recommendations, Assessment, Development and Evaluation (GRADE) approach to evaluate the quality of evidence for physical function and all-cause mortality through CINeMA web application[22]. Details of GRADE definition could be found in Appendix S7.

Results

Study Characteristics

After screening 22025 titles and abstracts and 237 full text articles, 49 studies including 5323 patients and 16 interventions, and 11 companion reports were included (Figure 1, Table 1 and Appendix S8). Forty-five studies were included in the meta-analysis. The details of remaining four studies are provided in Appendix S9. The most commonly included interventions were conventional open decompression (18 studies), midline preserving decompression (15 studies), epidural steroid with anaesthetic injection (11 studies), interspinous device (7 studies), conventional open decompression with fusion (8 studies), epidural steroid injection (6 studies), endoscopic decompression (6 studies) and conservative care (6 studies). Five or less studies were included for the remaining interventions.

The included studies were published between 1995 and 2019: 18 in Europe (37%), 16 in Asia/Oceania (33%), 13 in North America (26%) and two in Africa (4%). The majority of studies included a small sample (n<100, 30/49 studies, 61%) and a longer follow-up duration (>12 months, 31/49, 63%). Many studies did not report funding source (30/49, 61%), the level of stenosis (18/49, 37%), or type of stenosis (11/49, 22%).

Risk of Bias

Approximately half of the included studies presented high (21/45, 46%) overall risk of bias (Table 2, also see Appendix S10), with the main limitations including lack of blinding of participants, or those delivering the interventions (3/45, 7%); lack of blinding of outcome assessors (19/45, 42%); and lack of pre-specified analysis plan (6/45, 13%).

Primary Outcomes

Physical function

Twenty-six randomised controlled trials including 3247 patients and 13 interventions were included in the network meta-analysis for short-term physical function (Appendix S11)[23-47]. The

overall results showed no statistical difference between any surgical or invasive treatment and conservative care on physical function (Table 3, quality of evidence: moderate to very low, also see appendix S12). Twenty-seven randomised controlled trials including 3342 patients and nine (seven surgical interventions, one epidural injection and conservative care) interventions reported long-term assessments of physical function (Appendix S11)[24,25,33,35,39,40,42,44-59]. Once again, no statistical difference was observed when surgical or invasive treatments were compared with conservative care (Appendix S13, quality of evidence: low to very low, also see appendix S12).

All-cause mortality

Ten randomised controlled trials including 1573 patients and six (five surgical interventions and conservative care) interventions reported all-cause mortality (Appendix S11)[25,27,45,52,56,60-64].

Inconsistency test, sensitivity analysis and publication bias

The inconsistency test (Appendix S14) indicated minor inconsistency for physical function and no inconsistency for all-cause mortality. For both physical function and all-cause mortality, similar results were observed in the sensitivity analyses for both (Appendix S15), and we did not find publication bias from the comparison-adjusted funnel plot (Appendix S16).

Secondary Outcomes

For each outcome, the number of included studies, patients and interventions is listed in Appendix S17 and the network plot is listed in Appendix S11.

For back pain (short-term and long-term), long-term leg pain, mobility, and treatment withdrawal due to any reason, there was no statistical difference between any surgical treatment or invasive treatment and conservative care (Appendix S13).

For short-term leg pain, interspinous device (MD: -2.05, 95% CI: -3.98 to -0.12), midline splitting decompression (MD: -2.47, 95% CI: -4.45 to -0.5) and conventional open decompression (MD: -1.80, 95% CI: -3.49 to -0.11) were statistically superior to conservative care but the effects were too small to be clinically important.

For short-term overall pain, epidural steroid + hypertonic sodium + calcitonin injection (MD: -3.10, 95% CI: -5.56 to -0.64) was statistically superior to conservative care. The size of the effect was too small to be clinically important, however.

For short-term global impression of recovery, midline preserving decompression (OR: 4.76, 95% CI: 1.41 to 16.67) and endoscopic decompression (OR: 9.09, 95% CI: 1.75 to 50) were statistically superior to conservative care. For long-term global impression of recovery, midline preserving decompression was statistically superior to conservative care (OR: 9.09, 95% CI: 1.12 to 100).

For adverse effect due to any reason, conservative care (OR: 0.31, 95% CI: 0.10 to 0.94) was significantly associated with lower odds of adverse effect due to any reason compared to conventional open decompression with fusion.

For intervention related adverse effect, conservative care was significantly associated with lower odds of intervention related adverse effect compared to conventional open decompression with interspinous device (OR: 0.06, 95% CI: 0.01 to 0.33), midline splitting decompression (OR: 0.06, 95% CI: 0.01 to 0.39), conventional open decompression (OR: 0.05, 95% CI: 0.01 to 0.27), midline preserving decompression (OR: 0.15, 95% CI: 0.03 to 0.85), conventional open decompression with fusion (OR: 0.04, 95% CI: 0.01 to 0.24) and interspinous device only (OR: 0.18, 95% CI: 0.04 to 0.84).

For reoperation rate, conventional open decompression (OR: 0.22, 95% CI: 0.10 to 0.45), endoscopic decompression (OR: 0.13, 95% CI: 0.02 to 0.75), midline splitting decompression (OR: 0.24, 95% CI: 0.06 to 0.92), midline preserving decompression (OR: 0.10, 95% CI: 0.03 to 0.30) and conventional open decompression with fusion (OR: 0.22, 95% CI: 0.09 to 0.57) were all significantly associated with lower odds of reoperation rate compared to the interspinous device.

Ranking results

Details could be found in Appendix S18.

Discussion

This network meta-analysis incorporated 49 randomised controlled trials including 5323 patients comparing surgical and invasive procedures with conservative care in the treatment of degenerative LSS. The results showed that there is currently no evidence that a significant difference exists in physical function, mobility, or back pain between any surgical or invasive intervention and conservative care for the management of patients with degenerative LSS. Moreover, no statistically significant differences were shown between any treatment for reducing all-cause mortality. Interspinous device, midline splitting decompression and conventional open decompression were associated with a statistically significant but clinically unimportant improvement in short-term leg pain when compared with conservative care. With the exception of endoscopic decompression, all surgical interventions were associated with higher intervention related adverse effect when compared with conservative care. Of all available surgical procedures, interspinous device was associated with the highest rate of reoperation.

Strengths and limitations of this study

This systematic review has two main strengths. To our knowledge, it is the first comprehensive evidence synthesis to assess the effectiveness and safety of all available surgical and invasive interventions for the treatment of degenerative LSS based on published randomised controlled

trials. Comprehensive sensitivity analyses were performed to guarantee the robustness of the results.

A limitation of our study was that many of the included studies may have been affected by bias, such as English-language trials; insufficient blinding of participants; carers and treatment providers; inadequate blinding of outcome assessors; and lack of pre-published research protocols and/or data analysis plans. Another limitation was the timing of enrolment in included trials. Patients with extreme disability, neurologic progression, or simply inability to economically/financially endure prolonged impairment were excluded from the trials. We must exercise caution when generalizing the results to these kinds of patients. A final limitation of this review was the grouping of conservative care into one node. We, therefore, performed two additional sensitivity analysis (Appendix S5) and found no significant change to the results.

Comparison with other studies

The findings of this network meta-analysis are similar with the conclusions of two previous Cochrane reviews with similar topics, that there is insufficient evidence to support surgical management as a treatment of LSS [7,8,65]. Machado et al concluded a paucity of evidence on the efficacy of surgery for LSS[65]. In 2015, Zaina et al concluded that evidence was of insufficient quality to determine whether surgical treatment or a conservative approach is better for symptomatic (either neurogenic claudication or monoradicular or polyradicular symptoms) LSS[7]. Our findings on the lower rates of adverse effects and safety of conservative care when compared with surgical interventions was also consistent with a previous Cochrane review[7]. Furthermore our results corroborate with findings from one previous review which suggested that invasive interventions such as epidural steroids are not superior to conservative care and that epidural lidocaine with steroids was not found to be superior to lidocaine alone[6].

Patients with degenerative LSS are heterogeneous which might threaten the validity of the results, and previous studies paid less attention to it. To assess the potential influence from the

heterogeneity, we performed five sensitivity analyses based on the issues about typical symptom, spinal instability, spondylolisthesis, level affected, and inclusion criteria (Appendix S5). Because the surgical technologies might evolve without changing the name, we excluded studies which collected data before 2000 as an additional sensitivity analysis. Overall, the results from sensitivity analyses were similar with main analyses.

Implications for practice and research

This network meta-analysis provides the reader with the most comprehensive understanding of the current evidence for treatments of degenerative LSS. Readers may use the results of this review to assist in their recommendations and choices for treatment. Previous clinical practice guideline committees have had to rely on a small number of studies resulting in conflicting recommendations. For example, the North American Spine Society clinical guidelines recommend epidural steroid injections to provide short-term (two weeks to six months) symptom relief in patients with LSS and associated neurogenic claudication[1]. Our study showed that by assessing all available studies, there is no evidence to support epidural steroid injection above conservative care. Given the new evidence available from this network meta-analysis, updated clinical guidelines for the management of degenerative LSS is recommended. High quality cost-effectiveness analyses of surgical and invasive procedures for the management of degenerative LSS are also warranted to guide policy makers in their future recommendations. We also need future trials to clearly report the details of their conservative treatment protocols so that this part of studies could be reproduced.

Conclusions

There was no evidence to support that surgery or other invasive procedures are more effective or safer than conservative care at treating degenerative lumbar spinal stenosis.

References

1. Kreiner DS, Shaffer WO, Baisden JL, et al. An evidence-based clinical guideline for the diagnosis and treatment of degenerative lumbar spinal stenosis (update). *The spine journal : official journal of the North American Spine Society.* 2013;13(7):734-743.
2. Hartvigsen J, Hancock MJ, Kongsted A, et al. What low back pain is and why we need to pay attention. *Lancet (London, England).* 2018;391(10137):2356-2367.
3. Kalichman L, Cole R, Kim DH, et al. Spinal stenosis prevalence and association with symptoms: the Framingham Study. *The spine journal : official journal of the North American Spine Society.* 2009;9(7):545-550.
4. Deyo RA, Mirza SK, Martin BI, Kreuter W, Goodman DC, Jarvik JG. Trends, major medical complications, and charges associated with surgery for lumbar spinal stenosis in older adults. *Jama.* 2010;303(13):1259-1265.
5. Jansson KA, Blomqvist P, Granath F, Nemeth G. Spinal stenosis surgery in Sweden 1987-1999. *European spine journal : official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society.* 2003;12(5):535-541.
6. Manchikanti L, Knezevic NN, Boswell MV, Kaye AD, Hirsch JA. Epidural Injections for Lumbar Radiculopathy and Spinal Stenosis: A Comparative Systematic Review and Meta-Analysis. *Pain physician.* 2016;19(3):E365-410.
7. Zaina F, Tomkins-Lane C, Carragee E, Negrini S. Surgical versus non-surgical treatment for lumbar spinal stenosis. *The Cochrane database of systematic reviews.* 2016(1):Cd010264.
8. Overdevest GM, Jacobs W, Vleggeert-Lankamp C, Thome C, Gunzburg R, Peul W. Effectiveness of posterior decompression techniques compared with conventional laminectomy for lumbar stenosis. *The Cochrane database of systematic reviews.* 2015(3):Cd010036.
9. Dias S, Welton NJ, Sutton AJ, Ades AE. Evidence synthesis for decision making 1: introduction. *Medical decision making : an international journal of the Society for Medical Decision Making.* 2013;33(5):597-606.
10. Hutton B, Salanti G, Caldwell DM, et al. The PRISMA Extension Statement for Reporting of Systematic Reviews Incorporating Network Meta-analyses of Health Care Interventions: Checklist and ExplanationsPRISMA Extension for Network Meta-analysis. *Annals of Internal Medicine.* 2015;162(11):777-784.

11. Chen L, H Ferreira P, R Beckenkamp P, L Ferreira M. Comparative efficacy and safety of surgical and invasive treatments for adults with degenerative lumbar spinal stenosis: protocol for a network meta-analysis and systematic review. *BMJ Open.* 2019;9(4):e024752.

12. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ.* 2019;366:l4898.

13. Chaimani A, Salanti G. Visualizing assumptions and results in network meta-analysis: The network graphs package. *Stata Journal.* 2015;15(4):905-950.

14. Harris R, Bradburn M, Deeks J, Harbord R, Altman D, Sterne J. metan: fixed- and random-effects meta-analysis. *Stata Journal.* 2008;8(1):3-28.

15. White IR. Network meta-analysis. *Stata Journal.* 2015;15(4):951-985.

16. Chou R, Deyo R, Friedly J, et al. Systemic Pharmacologic Therapies for Low Back Pain: A Systematic Review for an American College of Physicians Clinical Practice Guideline. *Ann Intern Med.* 2017;166(7):480-492.

17. Ferreira ML, Herbert RD, Crowther MJ, Verhagen A, Sutton AJ. When is a further clinical trial justified? *Bmj.* 2012;345:e5913.

18. Higgins JP, Jackson D, Barrett JK, Lu G, Ades AE, White IR. Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. *Research synthesis methods.* 2012;3(2):98-110.

19. Dias S, Welton NJ, Caldwell DM, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. *Statistics in medicine.* 2010;29(7-8):932-944.

20. Salanti G, Ades AE, Ioannidis JP. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. *Journal of clinical epidemiology.* 2011;64(2):163-171.

21. Chaimani A, Higgins JP, Mavridis D, Spyridonos P, Salanti G. Graphical tools for network meta-analysis in STATA. *PloS one.* 2013;8(10):e76654.

22. Salanti G, Del Giovane C, Chaimani A, Caldwell DM, Higgins JP. Evaluating the quality of evidence from a network meta-analysis. *PloS one.* 2014;9(7):e99682.

23. Brown LL. A double-blind, randomized, prospective study of epidural steroid injection vs. the mild(R) procedure in patients with symptomatic lumbar spinal stenosis. *Pain practice : the official journal of World Institute of Pain.* 2012;12(5):333-341.

24. Celik SE, Celik S, Goksu K, Kara A, Ince I. Microdecompressive laminatomy with a 5-year follow-up period for severe lumbar spinal stenosis. *Journal of spinal disorders & techniques.* 2010;23(4):229-235.

25. Delitto A, Piva SR, Moore CG, et al. Surgery versus nonsurgical treatment of lumbar spinal stenosis: a randomized trial. *Ann Intern Med.* 2015;162(7):465-473.

26. Friedly JL, Comstock BA, Turner JA, et al. Long-Term Effects of Repeated Injections of Local Anesthetic With or Without Corticosteroid for Lumbar Spinal Stenosis: A Randomized Trial. *Archives of physical medicine and rehabilitation.* 2017;98(8):1499-1507.e1492.

27. Ghogawala Z, Dziura J, Butler WE, et al. Laminectomy plus Fusion versus Laminectomy Alone for Lumbar Spondylolisthesis. *The New England journal of medicine.* 2016;374(15):1424-1434.

28. Gurelik M, Bozkina C, Zafer K, Karadag O, Unal O, Bayrakli F. Unilateral laminotomy for decompression of lumbar stenosis is effective and safe: a prospective randomized comparative study. *Journal of Neurological Sciences (Turkish).* 2012;29(4):744-753.

29. Karm MH, Choi SS, Kim DH, et al. Percutaneous Epidural Adhesiolysis Using Inflatable Balloon Catheter and Balloon-less Catheter in Central Lumbar Spinal Stenosis with Neurogenic Claudication: A Randomized Controlled Trial. *Pain physician.* 2018;21(6):593-606.

30. Kim SH, Choi WJ, Suh JH, et al. Effects of transforaminal balloon treatment in patients with lumbar foraminal stenosis: a randomized, controlled, double-blind trial. *Pain physician.* 2013;16(3):213-224.

31. Koc Z, Ozcakir S, Sivrioglu K, Gurbet A, Kucukoglu S. Effectiveness of physical therapy and epidural steroid injections in lumbar spinal stenosis. *Spine.* 2009;34(10):985-989.

32. Koh WU, Choi SS, Park SY, et al. Transforaminal hypertonic saline for the treatment of lumbar lateral canal stenosis: a double-blinded, randomized, active-control trial. *Pain physician.* 2013;16(3):197-211.

33. Komp M, Hahn P, Oezdemir S, et al. Bilateral spinal decompression of lumbar central stenosis with the full-endoscopic interlaminar versus microsurgical laminotomy technique: a prospective, randomized, controlled study. *Pain physician.* 2015;18(1):61-70.

34. Lonne G, Johnsen LG, Rossvoll I, et al. Minimally invasive decompression versus x-stop in lumbar spinal stenosis: a randomized controlled multicenter study. *Spine.* 2015;40(2):77-85.

35. Malmivaara A, Slatis P, Heliovaara M, et al. Surgical or nonoperative treatment for lumbar spinal stenosis? A randomized controlled trial. *Spine.* 2007;32(1):1-8.

36. Manchikanti L, Cash KA, McManus CD, Damron KS, Pampati V, Falco FJ. A randomized, double-blind controlled trial of lumbar interlaminar epidural injections in central spinal stenosis: 2-year follow-up. *Pain physician.* 2015;18(1):79-92.

37. Manchikanti L, Cash KA, McManus CD, Pampati V, Fellows B. Fluoroscopic caudal epidural injections with or without steroids in managing pain of lumbar spinal stenosis: one-year results of randomized, double-blind, active-controlled trial. *Journal of spinal disorders & techniques.* 2012;25(4):226-234.

38. Manchikanti L, Cash KA, McManus CD, Pampati V, Singh V, Benyamin R. The preliminary results of a comparative effectiveness evaluation of adhesiolysis and caudal epidural injections in managing chronic low back pain secondary to spinal stenosis: a randomized, equivalence controlled trial. *Pain physician.* 2009;12(6):E341-354.

39. Moojen WA, Arts MP, Jacobs WC, et al. Interspinous process device versus standard conventional surgical decompression for lumbar spinal stenosis: randomized controlled trial. *Bmj.* 2013;347:f6415.

40. Musacchio MJ, Lauryssen C, Davis RJ, et al. Evaluation of Decompression and Interlaminar Stabilization Compared with Decompression and Fusion for the Treatment of Lumbar Spinal Stenosis: 5-year Follow-up of a Prospective, Randomized, Controlled Trial. *International journal of spine surgery.* 2016;10:6.

41. Nam H-S, Park YB. Effects of transforaminal injection for degenerative lumbar scoliosis combined with spinal stenosis. *Annals of rehabilitation medicine.* 2011;35(4):514.

42. Ruetten S, Komp M, Merk H, Godolias G. Surgical treatment for lumbar lateral recess stenosis with the full-endoscopic interlaminar approach versus conventional microsurgical technique: a prospective, randomized, controlled study. *Journal of neurosurgery Spine.* 2009;10(5):476-485.

43. Song SH, Ryu GH, Park JW, et al. The Effect and Safety of Steroid Injection in Lumbar Spinal Stenosis: With or Without Local Anesthetics. *Ann Rehabil Med.* 2016;40(1):14-20.

44. Stromqvist BH, Berg S, Gerdhem P, et al. X-stop versus decompressive surgery for lumbar neurogenic intermittent claudication: randomized controlled trial with 2-year follow-up. *Spine.* 2013;38(17):1436-1442.

45. Weinstein JN, Tosteson TD, Lurie JD, et al. Surgical versus nonsurgical therapy for lumbar spinal stenosis. *The New England journal of medicine.* 2008;358(8):794-810.

46. Yagi M, Okada E, Ninomiya K, Kihara M. Postoperative outcome after modified unilateral-approach microendoscopic midline decompression for degenerative spinal stenosis. *Journal of neurosurgery Spine.* 2009;10(4):293-299.

47. Zucherman JF, Hsu KY, Hartjen CA, et al. A multicenter, prospective, randomized trial evaluating the X STOP interspinous process decompression system for the treatment of neurogenic intermittent claudication: two-year follow-up results. *Spine.* 2005;30(12):1351-1358.

48. Anderson PA, Tribus CB, Kitchel SH. Treatment of neurogenic claudication by interspinous decompression: application of the X STOP device in patients with lumbar degenerative spondylolisthesis. *Journal of neurosurgery Spine.* 2006;4(6):463-471.

49. Azzazi A, Elhawary Y. Dynamic stabilization using X-stop versus transpedicular screw fixation in the treatment of lumbar canal stenosis; comparative study of the clinical outcome. *Neurosurgery Quarterly.* 2010;20(3):165-169.

50. Benyamin RM, Staats PS, Mi DASEI. MILD(R) Is an Effective Treatment for Lumbar Spinal Stenosis with Neurogenic Claudication: MiDAS ENCORE Randomized Controlled Trial. *Pain physician.* 2016;19(4):229-242.

51. Cho DY, Lin HL, Lee WY, Lee HC. Split-spinous process laminotomy and discectomy for degenerative lumbar spinal stenosis: a preliminary report. *Journal of neurosurgery Spine.* 2007;6(3):229-239.

52. Forsth P, Olafsson G, Carlsson T, et al. A Randomized, Controlled Trial of Fusion Surgery for Lumbar Spinal Stenosis. *The New England journal of medicine.* 2016;374(15):1413-1423.

53. Marsh GD, Mahir S, Leyte A. A prospective randomised controlled trial to assess the efficacy of dynamic stabilisation of the lumbar spine with the Wallis ligament. *European spine journal : official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society.* 2014;23(10):2156-2160.

54. Mobbs RJ, Li J, Sivabalan P, Raley D, Rao PJ. Outcomes after decompressive laminectomy for lumbar spinal stenosis: comparison between minimally invasive unilateral laminectomy for bilateral decompression and open laminectomy: clinical article. *Journal of neurosurgery Spine.* 2014;21(2):179-186.

55. Rajasekaran S, Thomas A, Kanna RM, Prasad Shetty A. Lumbar spinous process splitting decompression provides equivalent outcomes to conventional midline decompression in degenerative lumbar canal stenosis: a prospective, randomized controlled study of 51 patients. *Spine.* 2013;38(20):1737-1743.

56. Schmidt S, Franke J, Rauschmann M, Adelt D, Bonsanto MM, Sola S. Prospective, randomized, multicenter study with 2-year follow-up to compare the performance of decompression with and without interlaminar stabilization. *Journal of neurosurgery Spine.* 2018;28(4):406-415.

57. Skoro I, Stancic M, Kovacevic M, Duric KS. Long-Term Results and Efficacy of Laminectomy with Fusion Versus Young Laminoplasty for the Treatment of Degenerative Spinal Stenosis. *World neurosurgery.* 2016;89:387-392.

58. Thome C, Zevgaridis D, Leheta O, et al. Outcome after less-invasive decompression of lumbar spinal stenosis: a randomized comparison of unilateral laminotomy, bilateral laminotomy, and laminectomy. *Journal of neurosurgery Spine.* 2005;3(2):129-141.

59. Watanabe K, Matsumoto M, Ikegami T, et al. Reduced postoperative wound pain after lumbar spinous process-splitting laminectomy for lumbar canal stenosis: a randomized controlled study. *Journal of neurosurgery Spine.* 2011;14(1):51-58.

60. Amundsen T, Weber H, Nordal HJ, Magnaes B, Abdelnoor M, Lilleas F. Lumbar spinal stenosis: conservative or surgical management?: A prospective 10-year study. *Spine.* 2000;25(11):1424-1435; discussion 1435-1426.

61. Haddadi K, Ganjeh Qazvini HR. Outcome after Surgery of Lumbar Spinal Stenosis: A Randomized Comparison of Bilateral Laminotomy, Trumpet Laminectomy, and Conventional Laminectomy. *Frontiers in surgery.* 2016;3:19.

62. Ko S, Oh T. Comparison of bilateral decompression via unilateral laminotomy and conventional laminectomy for single-level degenerative lumbar spinal stenosis regarding low back pain, functional outcome, and quality of life - A Randomized Controlled, Prospective Trial. *Journal of orthopaedic surgery and research.* 2019;14(1):252.

63. Slatis P, Malmivaara A, Heliovaara M, et al. Long-term results of surgery for lumbar spinal stenosis: a randomised controlled trial. *European spine journal : official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society.* 2011;20(7):1174-1181.

64. Weinstein JN, Lurie JD, Tosteson TD, et al. Surgical versus nonsurgical treatment for lumbar degenerative spondylolisthesis. *The New England journal of medicine.* 2007;356(22):2257-2270.

65. Machado GC, Ferreira PH, Yoo RI, et al. Surgical options for lumbar spinal stenosis. *The Cochrane database of systematic reviews.* 2016;11:Cd012421.

Table 1. Characteristics of the 49 included studies.

Characteristics	No (%) of studies
Publication year:	
1995-2000	3 (6)
2001-2006	3 (6)
2007-2012	17 (35)
2013-2019	26 (53)
Geographical region:	
Asia/Oceania	16 (33)
Europe	18 (37)
North America	13 (26)
Africa	2 (4)
Study setting:	
Single centre	21 (43)
Multicentre	15 (31)
Not reported	13 (26)
Study duration (months):	
0-12	18 (37)
13-24	19 (39)
25-36	0 (0)
37-48	4 (8)
49-60	3 (6)
>60	5 (10)
Funding source:	
Commercial	7 (14)
Government	9 (18)
Hospital	3 (6)
Not reported	30 (61)
Outcomes:	
Physical function	46 (94)
All-cause mortality	21 (43)
Leg pain	22 (45)
Back pain	23 (47)
Overall pain	13 (27)
Global impression of recovery	18 (37)
Mobility	10 (20)
Adverse effect due to any reason	27 (55)
Intervention related adverse effect	26 (53)
Reoperation rate	12 (24)
Treatment withdraw due to any reason	28 (57)
Age group (years, mean):	
<65	22 (45)
≥65	25 (51)
Not reported	2 (4)

Proportion of men (%):	
<50	27 (55)
≥50	21 (43)
Not reported	1 (2)
Sample size:	
<100	30 (61)
≥100	19 (39)
Stenosis level:	
Single	4 (8)
Multiple	27 (55)
Not reported	18 (37)
Stenosis type:	
Central	28 (57)
Lateral	1 (2)
Foraminal	1 (2)
Mixed	8 (16)
Not reported	11 (22)
Interventions:	
CD	18 (37)
ID	8 (16)
CD+ID	3 (6)
CD+Fu	8 (16)
Endo	6 (12)
MPD	15 (31)
MSD	5 (10)
Epi	1 (2)
EpiS	6 (12)
BT	2 (4)
EpiSH	1 (2)
EpiA	5 (10)
EpiAS	11 (22)
EpiASC	1 (2)
EpiASH	1 (2)
Cons	6 (12)

CD: conventional open decompression; ID: interspinous device; CD+ID: conventional open decompression with interspinous device; CD+Fu: conventional open decompression with fusion; MSD: midline splitting decompression; MPD: midline preserving decompression; Endo: endoscopic decompression; EpiS: epidural steroid injection; EpiA: epidural anaesthetic injection; EpiAS: epidural steroid + anaesthetic injection; BT: balloon treatment with epidural injection; EpiASH: epidural steroid + anaesthetic + hypertonic sodium injection; EpiASC: epidural steroid + hypertonic sodium + calcitonin injection; Cons: conservative care; Epi: epidural injection with saline solution only; EpiSH: epidural steroid + hypertonic sodium injection

Table 2. Summary of risk of bias assessment (n= 45 randomised controlled trials). Values are numbers (percentages).

Assessment item	Risk level		
	Low	Some concerns	High
Arising from the randomization process	21 (47)	24 (53)	0 (0)
Due to deviations from the intended interventions (effect of assignment to intervention)	14 (31)	28 (62)	3 (7)
Missing outcome data	39 (87)	0 (0)	6 (13)
Measurement of the outcome	24 (53)	2 (4)	19 (42)
Selection of the reported result	20 (44)	25 (56)	0 (0)
Overall risk of bias	7 (16)	17 (38)	21 (46)

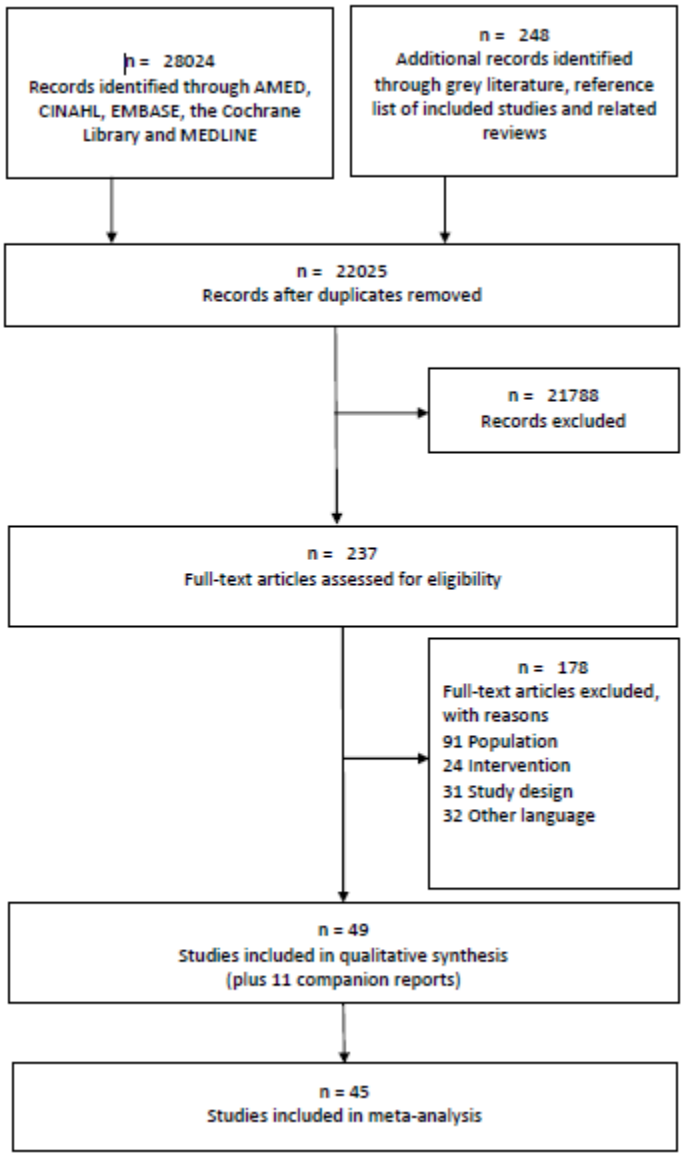
Table 3. Results from pairwise and network meta-analyses for short-term physical function (The numbers are presented as standardised mean difference and 95% confidence interval).

Cons					-0.4 (-1.92,1.12)			0.19 (-0.6,0.97)				
0.89 (-0.14,1.92)	ID						**-1.55 (-2.84,-0.26)**	0.31 (-0.21,0.83)	**-2.46 (-3.69,-1.23)**			
0.49 (-1.21,2.19)	-0.40 (-2.20,1.40)	EpiASH	0.73 (-0.55,2)		**-2.15 (-3.52,-0.78)**							
0.46 (-1.22,2.14)	-0.43 (-2.21,1.35)	-0.03 (-1.19,1.13)	BT		-0.51 (-1.79,0.76)							
0.06 (-1.37,1.49)	-0.83 (-2.23,0.57)	-0.43 (-2.53,1.67)	-0.40 (-2.48,1.68)	MSD			0.37 (-1.06,1.8)	0.16 (-1.26,1.59)				
-0.81 (-2.03,0.41)	**-1.70 (-3.05,-0.35)**	**-1.30 (-2.49,-0.11)**	**-1.27 (-2.42,-0.11)**	-0.87 (-2.59,0.86)	EpiAS	0.15 (-0.54,0.83)			0.49 (-1.01,2)			
-0.66 (-2.06,0.73)	**-1.55 (-3.06,-0.04)**	-1.15 (-2.52,0.22)	-1.12 (-2.46,0.22)	-0.72 (-2.58,1.14)	0.15 (-0.54,0.83)	EpiA						
0.36 (-0.59,1.30)	-0.53 (-1.35,0.29)	-0.13 (-1.91,1.64)	-0.10 (-1.85,1.65)	0.30 (-0.94,1.53)	1.16 (-0.15,2.48)	1.02 (-0.46,2.50)	MPD	-0.45 (-1,0.1)		0 (-0.94,0.04)		
0.29 (-0.45,1.03)	-0.60 (-1.36,0.17)	-0.20 (-1.93,1.54)	-0.17 (-1.88,1.55)	0.23 (-1.00,1.47)	1.10 (-0.16,2.36)	0.95 (-0.48,2.39)	-0.06 (-0.69,0.56)	CD			0.32 (-1.08, 1.72)	
-0.72 (-1.87,0.43)	**-1.61 (-2.60,-0.62)**	-1.21 (-2.91,0.49)	-1.18 (-2.85,0.50)	-0.78 (-2.30,0.74)	0.09 (-1.12,1.30)	-0.06 (-1.45,1.33)	**-1.08 (-2.05,-0.10)**	**-1.01 (-2.02,-0.00)**	EpiS	0.54 (-0.43, 1.51)		
0.10 (-1.05,1.25)	-0.79 (-1.81,0.23)	-0.39 (-2.19,1.41)	-0.36 (-2.14,1.42)	0.04 (-1.42,1.50)	0.91 (-0.44,2.26)	0.76 (-0.76,2.28)	-0.26 (-1.09,0.57)	-0.19 (-1.16,0.77)	0.82 (-0.03,1.67)	Endo		

0.61 (-0.98,2.19)	-0.28 (-1.87,1.32)	0.12 (-2.11,2.35)	0.15 (-2.06,2.36)	0.55 (-1.31,2.42)	1.42 (-0.47,3.30)	1.27 (-0.74,3.28)	0.25 (-1.28,1.79)	0.32 (-1.08,1.72)	1.33 (-0.40,3.05)	0.51 (-1.19,2.21)	CD+Fu	0.21 (-0.89,1.32)
0.82 (-1.24,2.89)	-0.07 (-2.14,2.01)	0.33 (-2.26,2.93)	0.37 (-2.21,2.94)	0.77 (-1.52,3.06)	1.63 (-0.67,3.94)	1.48 (-0.92,3.89)	0.47 (-1.56,2.50)	0.53 (-1.40,2.46)	1.54 (-0.63,3.72)	0.72 (-1.43,2.88)	0.21 (-1.11,1.54)	CD+ID

The results from bottom-left are network evidence. The results from top-right are pairwise evidence. For the network evidence, the minus sign means the up to the left intervention has the superior result. For the pairwise evidence, the minus sign means the leftward intervention has the superior result. The statistically significant results were presented in bold. CD: conventional open decompression; ID: interspinous device; CD+ID: conventional open decompression with interspinous device; CD+Fu: conventional open decompression with fusion; MSD: midline splitting decompression; MPD: midline preserving decompression; Endo: endoscopic decompression; EpiS: epidural steroid injection; EpiA: epidural anaesthetic injection; EpiAS: epidural steroid + anaesthetic injection; BT: balloon treatment with epidural injection; EpiASH: epidural steroid + anaesthetic + hypertonic sodium injection; Cons: conservative care.

Figure 1. Flow diagram.



Appendix

Abbreviations

AE: adverse effect

BP: back pain

BT: balloon treatment with epidural injection

C: central

CD: conventional open decompression

CD+Fu: conventional open decompression with fusion

CD+ID: conventional open decompression with interspinous device

DE: death

DS: degenerative spondylolisthesis

Endo: endoscopic decompression

EpiA: epidural anaesthetic injection

EpiAS: epidural steroid + anaesthetic injection

EpiASC: epidural steroid + hypertonic sodium + calcitonin injection

EpiASH: epidural steroid + anaesthetic + hypertonic sodium injection

EpiS: epidural steroid injection

F: foraminal

GIR: global impression of recovery

HG: high grade: >50% or indicated as require surgical interventions

ID: interspinous device

L: lateral

LG: low grade

LP: leg pain

MB: mobility

MPD: midline preserving decompression

MSD: midline splitting decompression

NR: not reported

NRS: Numeric Rating Scale

ODI: Oswestry Disability Index

OP: overall pain

PF: physical function

QoL: quality of life

RMDQ: Roland Morris Disability Questionnaire

SMD: Standardized Mean Difference.

TW: treatment withdraw

VAS: Visual Analog Scale

Appendix S1: Characteristics of conservative care

First author, Year	Description
Amundsen, 2000	The patients were fitted with a 3-point brace (a hyperextension thoracolumbar orthosis with pelvic support/Camp) and transferred to the rehabilitation department for 1 month.
Delitto, 2015	Lumbar flexion exercises, general conditioning exercises, and patient education.
Koc, 2009	Group 1: ultrasound 1.5 W/cm^2 for 10 minutes (Sonostat 633 model, Siemens), hot pack for 20 minutes, and TENS (Bio Tens ST-606 M model) for 20 minutes to the lumbar region. Group 3: No treatment.
Malmivaara, 2007	The patients were prescribed nonsteroidal anti-inflammatory drugs when indicated and were referred to Physiotherapists.
Weinstein, 2007	at least active physical therapy, education or counseling including instructions for exercising at home, and nonsteroidal anti-inflammatory agents if tolerated
Weinstein, 2008	At least active physical therapy, education or counselling with home exercise instruction, and the administration of nonsteroidal anti-inflammatory drugs, if tolerated.

Appendix S2: Definition for adverse effect outcomes

Adverse effect due to any reason

All mentioned adverse effects except those defined as reoperation.

Intervention related adverse effect

We classified adverse effects into this category if the study used the terms ‘intervention related’ or ‘related synonyms’.

Reoperation rate

To make this outcome more comparable, we only included studies which reported 2-year data for reoperation rate. Because cross-over from the conservative care group to surgical group is different from reoperation in surgical group, we excluded studies with a conservative care group in the analysis.

Appendix S3: Magnitude of Effects on pain intensity and physical function

Pain intensity

Rating scale	Slight/Small	Moderate	Large/Substantial
	If the studies use the same rating scale		
VAS (0-100)	5-10	11-20	>20
NRS (0-10)	0.5-1	1.1-2	>2
	If the studies use the different rating scales		
SMD	0.2-0.5	0.6-0.8	>0.8

Physical function

Rating scale	Slight/Small	Moderate	Large/Substantial
	If the studies use the same rating scale		
ODI (0-100)	5-10	11-20	>20
RMDQ (0-24)	1-2	3-5	>5
	If the studies use the different rating scales		
SMD	0.2-0.5	0.6-0.8	>0.8

Appendix S4: Inclusion and exclusion criteria

First author, Year	Inclusion criteria	Exclusion criteria
Amundsen 2000	sciatic pain in the leg(s), with or without pain in the back, together with radiologic signs of stenosis and compression of the clinically afflicted nerve root(s).	a bulging or herniated disc, spondylolysis, coxarthosis, gonarthosis, arterial insufficiency in the legs, polyneuropathy, concomitant serious disease, or previous surgery on the back.
Anderson 2006	at least 50 years of age, had to have their symptoms relieved by sitting or flexion, and had to have completed at least a 6-month course of nonoperative treatment.	could not walk at least 50 feet and/or were unable to sit for at least 50 minutes, or if anterior translation greater than 25% was seen on imaging studies.
Azzazi 2010	degenerative spondylolisthesis or retrolisthesis (up to Grade I), to have lateral and/or central spinal stenosis as diagnosed on neuroimaging studies, and to qualify for posterior lumbar spinal decompression and instrumented fusion for single-level or contiguous two-level disease between L-1 and S-1.	undergone an earlier lumbar fusion or decompression attempts, earlier total facetectomy, or trauma at the affected level.
		with diseases that preclude surgical management (severe osteoporosis, osteopenia, immune suppression, malignancy, and active local and/or systemic infection).
	a predominant component of leg pain (a preoperative score of 40mm on a 100-mm VAS) rather than back pain symptoms.	younger than 20 years or older than 80 years of age or those with morbid obesity as measured by a BMI greater than 40.
	at least moderate disability and were unresponsive to conservative management for a minimum of 3 months.	
Benyamin 2016	≥ 65 years old, a Medicare beneficiary, and have had neurogenic claudication symptoms for at least 3 months that was refractory to physical therapy, home exercise programs, and oral analgesics.	with an Oswestry Disability Index (ODI) score < 31 or Numeric Pain Rating Scale (NPRS) score < 5.
		with a history of spinal fracture with current related pain, prior surgery at any treatment level, or motor deficit or disabling back or leg pain from causes other than LSS neurogenic claudication.
	LSS with ligamentum flavum > 2.5mm was confirmed by preoperative magnetic resonance imaging or computed tomography.	with Grade III or higher spondylolisthesis, and those suffering from epidural lipomatosis, if deemed to be a significant contributor of canal narrowing by the physician.
	underwent predefined and precise diagnostic screening to confirm symptoms of neurogenic claudication prior to enrollment in the study.	past epidural injection therapy was not an exclusion criteria.
	with lumbar spine comorbid conditions commonly associated with spinal stenosis, including osteophytes, facet hypertrophy, minor spondylolisthesis, foraminal stenosis, and disc protrusion, were included unless the treating physician determined that the condition was too advanced.	unable to walk ≥ 10 feet unaided before being limited by pain.
Brown 2012	symptomatic LSS patients with painful lower limb neurogenic claudication and	they had prior surgery at the intended treatment level or had previously been treated with epidural steroids.

	hypertrophic ligamentum flavum as a contributing factor. at least 18 years of age, had previously failed conservative therapy, and presented with an Oswestry Disability Index (ODI) score > 20. radiologic evidence showed evidence of LSS (L3–L5), ligamentum flavum >2.5 mm was confirmed by preoperative MRI or CT, central canal cross sectional area was £ 100 mm, and anterior listhesis was confirmed to be £ 5.0 mm for all patients. all patients were able to walk at least 10 feet unaided before being limited by pain and provided written informed consent.	a history of recent spinal fractures, disabling back or leg pain from causes other than LSS, fixed spondylolisthesis > Grade 1, disk protrusion or osteophyte formation, or excessive facet hypertrophy bleeding disorders, current use of anticoagulants, or wound healing pathologies deemed to compromise outcomes, such as diabetes, cancer, severe COPD, or those that had used ASA or NSAID within 5 days of treatment pregnant or breastfeeding, unable to lie prone for any reason with anesthesia support, unable to give informed consent, on Workman's Compensation or considering litigation associated with back pain
Celik 2010	severe back/leg pain and neurogenic claudication (with different degrees of muscle weakness according to the stenotic level). had not responded to conservative medical therapy and a physical therapy program. all the patients were showed more than 41% in ODI, more pain than 7 in VAS with walking distance less than 30 meters. The patients were classified as severe lumbar spinal stenosis clinically. all the radiologic studies were studied to achieve definitive preoperative diagnosis; dynamic x-rays, thin-sliced computerized tomography (CT) and magnetic resonance images (MRIs).	the patients requiring discectomy or showing any kind of instability before the surgery were also excluded.
Cho 2007	evidence of spinal stenosis was obtained from CT scans or MR images. spinal stenosis was defined by an anteroposterior diameter of the spinal canal less than 11 mm, an interpediculate distance of less than 16 mm, and a lateral recess distance of less than 3 mm hypertrophic facets and ligamentum flavum, and a bulging disc were typically found. the usual clinical symptoms were lumbago and intermittent claudication. conservative treatments, including medication, rehabilitation, rest, or wearing	elderly patients (more than 80 years of age) with higher anesthetic risks or severe medical comorbidities, such as congestive heart failure, uremia, liver cirrhosis, coagulopathy, and others, were excluded, as were patients with lumbar stenosis and spondylolisthesis requiring additional instrumentation.

	a brace were attempted for at least 6 months before surgery.	
Delitto 2015	with a diagnosis of LSS identified by computed tomography using the criteria of Wiesel and colleagues or magnetic resonance imaging using the criteria of Boden and colleagues. all patients were considered by a spine surgeon to be candidates for surgical decompression and had consented to surgery. presence of neurogenic claudication (for example, self-reported inability to walk more than a quarter mile because of lower-extremity pain or cramping); consent to be randomly assigned to surgery or a specified PT clinic for twice-weekly exercise sessions; and no previous surgery for LSS at the level being considered for decompression.	younger than 50 years, had signs of serious dementia, were diagnosed with severe vascular disease or had a recent history of myocardial infarction, had concomitant spondylolisthesis requiring spinal fusion (defined as >5 mm of slippage), had compression fractures at the level being considered for decompression, or were diagnosed with metastatic cancer.
Elsheikh 2016	over 40 years old with a history of chronic low back pain with or without lower extremity pain ≥ 6 on a visual analog scale (VAS) of 0 – 10; pain for at least 3 months; with a diagnosis of central spinal stenosis with or without radicular pain (confirmed by computed tomography [CT] revealed anterior-posterior diameter < 12 mm at the level of the lumbar vertebrae) failed to improve with conservative management, including physical therapy, exercises, and pharmacotherapy	INR > 1.5; platelet count < 50,000; infection at the site of needle entry; congenital spinal canal stenosis; degenerative spondylolithesis, psychiatric disorders affecting co-operation of the patient, a history of spine surgery, previous chronic opioid use, peripheral vascular disease, uncontrolled medical illness (diabetes and/or hypertension), and patients with a history of adverse reaction to either local anesthetics, steroids, or calcitonin
Forsth 2016	pseudoclaudication in one or both legs and back pain (score on visual-analogue scale >30) 1 or 2 adjacent stenotic segments (cross-section area of the dural sac ≤75 mm^2) between L2 and the sacrum on magnetic resonance imaging duration of symptoms >6 mo between 50 and 80 years of age who had received a diagnosis of lumbar spinal stenosis	spondylolysis degenerative lumbar scoliosis (Cobb angle >20 degrees) history of lumbar spinal surgery for spinal stenosis or instability stenosis not caused by degenerative changes stenosis caused by a herniated disk other specific spinal conditions (e.g., ankylosing spondylitis, cancer, or neurologic disorders) history of vertebral compression fractures in affected segments

		psychological disorders (e.g., dementia or drug abuse) that caused the surgeon to consider participation to be inappropriate
Friedly 2016	pain in the low back, buttock, and/or lower extremity (rating of average pain in past week>4 on 0-10 scale) with standing, walking and/or spinal extension (buttock/leg>back pain). Roland-Morris Disability Questionnaire (modified to specify back or leg pain-related problems) score of at least 7. mild-moderate-severe lumbar central canal spinal stenosis identified by MRI or CT according to the criteria of Boden et al. lower extremity symptoms consistent with neurogenic claudication. age 50 years or older	fibromyalgia diagnosis, chronic widespread pain, lower extremity amputation, Parkinson's, head injury, stroke, other neurologic conditions severe vascular, pulmonary or coronary artery disease that limits ambulation including recent myocardial infarction (within the last 6 months) spinal instability requiring surgery severe osteoporosis as defined by multiple compression fractures or a fracture at the same level as the stenosis metastatic cancer excessive alcohol consumption or evidence of non-prescribed or illegal drug use as determined by the two-item conjoint screen (TICS) screening questionnaire (1 or more positive answer) possible pregnancy or other condition that precludes the use of fluoroscopy concordant pain with internal rotation of the hip (or known hip joint pathology) active local or systemic infection allergy to local anesthetic, steroid or contrast
Ghogawala 2016	patients between age 50 and 75 with symptomatic lumbar spinal stenosis and single level grade I degenerative spondylolisthesis (3 –14 mm). symptomatic spinal stenosis will be defined as radicular and/or back pain either induced or aggravated by activity and relieved by rest in a patient with either moderately severe or severe spinal stenosis as determined by an independent radiologist.	gross spinal instability defined as movement greater than 3 mm on flexion/ extension studies history of previous lumbar spinal surgery serious medical illness (ASA Class III or higher) spondylolysis, multilevel spondylolisthesis, or high grade spondylolisthesis (Grade II or greater than 14 mm)
Grob 1995	the spinal stenosis was diagnosed on the basis of the history and clinical examination as well as computerized tomographic scans that had been made after myelography or magnetic resonance imaging studies. Specimens of the cerebrospinal fluid were examined to exclude the possiblity of systemic disease. Anteroposterior and lateral roentgenograms were made of all patients	patients who had obvious instability of the lumbar spine were not included in this study. Instability was diagnosed on the basis of (1) a concomitant slip of a vertebra of more than five millimeters or another gross deformity such as rotational instability characterized by more than five millimeters of lateral offset on the anteroposterior roentgenogram or degenerative scoliosis. (2) spondylolysis with an osseous defect of the pars interarticularis, or (3) a previous operation on the lumbar spine.

	to evaluate the degree of stenosis; a midsagittal diameter of the spinal canal of less than eleven millimeters was considered stenotic.	
Gurelik 2012	symptoms of neurogenic claudication or radiculopathy	associated pathological entities such as instability and significant disc herniation
	radiological evidence of degenerative lumbar stenosis	previous surgery for lumbar spine disorder
Haddadi 2016	indications of neurogenic claudication or radiculopathy	exhibiting stable spondylolisthesis or having a past of surgery for herniated lumbar discs
	neuroimaging signs of degenerative stenosis	diabetic patients and osteoporotic or heavy smoker patients
	lack of related pathological matters such as disc herniation or instability	
	no presence of surgery for lumbar stenosis or fusion	
	symptoms were measured as intractable to non-surgical organization if traditional trials, principally non-steroidal anti-inflammatory drugs and somatic therapies, had been used for at least 12 weeks without enough improvement	
	back and leg visual analog scale (VAS) above score seven	
Kang 2019	clinically and radiologically diagnosed with central spinal stenosis in the lumbar spine	previous spine surgery, infection, trauma, and tumors
	neurogenic claudication, unresponsive to conservative treatment, and single-level central canal stenosis without evidence of instability	
	all patients had undergone nonoperative treatment for at least 3 months before surgery. Patients were recommended for surgery if they had failed nonoperative treatment and continued to have significant pain and daily activity restrictions due to neurogenic claudication or radicular pain	
Karm 2018	chronic LSS patients aged ≥ 40 years	chronic LSS patients aged ≥ 40 years
	lower back pain and/or lumbar radicular pain intensity ≥ 6 (out of 10) on the Numerical Rating Scale (NRS-11), and neurogenic intermittent claudication	lower back pain and/or lumbar radicular pain intensity ≥ 6 (out of 10) on the Numerical Rating Scale (NRS-11), and neurogenic intermittent claudication
	confirmed diagnosis of moderate or severe central, but not foraminal or lateral recess,	confirmed diagnosis of moderate or severe central, but not foraminal or lateral recess, LSS by magnetic resonance imaging (MRI)

	LSS by magnetic resonance imaging (MRI)	previous failure of conservative management, such as exercise therapy, physical therapy, or analgesic medication
	previous failure of conservative management, such as exercise therapy, physical therapy, or analgesic medication	
Kim 2013	patients with unilateral radicular pain with positive provocation factors that were not relieved by routine conservative treatments consisting of physiotherapy, exercise, analgesic medications, and epidural steroid injection for at least 6 months	acute back or leg pain; patients who developed signs of progressive neurologic deficits, including muscle atrophy and abnormal tendon reflexes; and patients with a history of prior spine surgery, allergic response to steroid or contrast dye, and bleeding diathesis or overt coagulopathy
	positive provocation factors included leg symptoms elicited or aggravated by walking but relieved by sitting down	bilateral radiculopathy or spinal stenosis at more than 3 levels
	a thorough history and physical examination was performed to rule out the confounding diagnosis of vascular disease or other origins.	
Ko 2019	patients with degenerative lumbar spinal stenosis requiring surgery due to neurogenic claudication with radiculopathy	patients who underwent spinal surgery in the past
		in addition to decompression, patients who needed further segmental fusion surgery.
	patients with one-level central stenosis requiring decompression	patients who require multiple segments of decompression surgery (≥ 3 levels)
	patients with MRI findings consistent with symptoms on preoperative radiological examination	patients with cervical lesions other than lumbar lesions
		patient with rapidly progressive neurological deficit
		patients who cannot cooperate in completing the questionnaire due to dementia or stroke
		neuromuscular disorder
		spinal malignancy, spinal infection, etc.
Koc 2009	diagnosed as LSS by medical history, physical and neurologic examination, as well as MRI findings	with the history of coronary artery disease, peripheral artery disease, spinal surgery, recent vertebral fracture, progressive neurologic deficit, or cauda equina syndrome
Komp 2015	predominant leg symptoms, neurogenic claudication with or without paresis, back pain maximum 30/100 on the visual analogue scale (VAS), conservative therapy exhausted or no longer indicated due to the symptoms, monosegmental central stenosis caused by facet hypertrophy, hypertrophy of the ligamentum flavum, and disc protrusions or the combination of those	predominant back pain, foraminal stenosis in the lower level, fresh soft disc herniations with bony stenosis, degenerative spondylolisthesis more than Meyerding Grade I, multidirectional rotation slide, scoliosis more than 20°, prior surgery in the same segment, and cauda equina syndrome
Liu 2013	LSS without degenerative spondylolisthesis or interbody instability	

Lonne 2015	aged 50 to 85 years, exhibited symptoms of neurogenic intermittent claudication within 250-m walking distance for at least 6 months, and were treated conservatively without sufficient effect or such treatment was considered as inexpedient the relief of symptoms through spinal flexion was an inclusion criterion. If in doubt, patients were asked in detail about situations that provided relief, where flexion relief was considered if 2 of the following conditions were present: the patient was able to sit for more than 30 minutes without pain, walk longer with a walking aid, bicycle a long distance without pain, and/or used to sleep in a flexed position to avoid pain. patients with 1 or 2 stenotic levels (from L2 to L5) and with minor spondylolisthesis (Meyerding, grade 1)	all participants had preoperative magnetic resonance images and radiographs of the lower spine to rule out osteoporotic fractures, deformity, or signs of instability
Slatis 2011	1) clinical symptoms: back pain radiation to lower limbs or buttocks; fatigue or loss of sensation in the lower limbs aggravated by walking. 2) persistent pain without progressive neurologic dysfunction. 3) imaging techniques: spinal canal narrowing, the sagittal diameter of the dural sac being less than 10 mm^2, or the planimetrically assessed cross-sectional dural area being less than 75 mm^2. 4) duration of symptoms and signs for more than 6 months. 5) clinical signs and symptoms corresponding to segmental radiographic level of stenosis. 6) severity of the disease justifying either surgical or nonoperative treatment The following conditions did not prevent inclusion: radiographic instability of the lumbar spine; degenerative spondylolisthesis; sick leave or early retirement because of degenerative LSS; mild motor or sensory impairment in the lower limbs; well-functioning hip or knee prosthesis	1) severe LSS with intractable pain and progressive neurologic dysfunction, suggesting forthcoming surgical treatment 2) mild LSS, characterized by radiographic narrowing of the lumbar spinal canal, but clinical signs and symptoms feeble enough to exclude surgical intervention 3) spinal stenosis not caused by degeneration, e.g., congenital spinal stenosis 4) spondylolysis and spondylolytic spondylolisthesis 5) an earlier back operation because of spinal stenosis or instability 6) lumbar herniated disc diagnosed during the last 12 months 7) another specific spinal disorder, e.g., ankylosing spondylitis, neoplasm, or metabolic diseases 8) intermittent claudication due to atherosclerosis 9) severe osteoarthrosis or arthritis causing dysfunction of the lower limbs 10) neurologic disease causing impaired function of the lower limbs, including diabetic neuropathy 11) psychiatric disorders 12) alcoholism
Manchikanti 2009	diagnosis of lumbar central spinal stenosis with radicular pain, patients over the age of 50 years; patients with a history of chronic function-limiting low back pain and lower extremity pain of at least 6 months duration	history of lumbar surgery, central spinal stenosis without radicular pain, foraminal stenosis, uncontrollable or unstable opioid use, uncontrolled psychiatric disorders, uncontrolled medical illness either acute or chronic, any conditions that could interfere with the interpretation of the outcome assessments, pregnant or lactating women, and

	patients who have failed to improve substantially with conservative management including, but not limited to, physical therapy, chiropractic manipulation, exercises, drug therapy, and bed rest. All these patients had also failed fluoroscopically directed epidural injections.	patients with a history or potential for adverse reaction(s) to local anesthetics or steroids
Manchikanti 2012	patients with central spinal stenosis with radicular pain of at least 6 months duration pain must have been function-limiting, 30 years or older failed conservative management	patients with a history of uncontrollable or unstable opioid use, uncontrolled psychiatric disorders, uncontrolled medical illness, those suffering with conditions that could interfere with the interpretation of outcome assessments, pregnant or lactating women, and those with a history or potential for adverse reactions to lidocaine or betamethasone
Manchikanti 2015	patients with central spinal stenosis with radicular pain of at least 6 months duration at least 30 years of age with a history of chronic function-limiting low back and lower extremity pain of at least 6 months duration all patients must have undergone conservative management with insufficient improvement	foraminal stenosis without central spinal stenosis, previous history of surgery, and uncontrollable or unstable psychiatric disorders, medical disorders, or opioid use any conditions that could interfere with the interpretation of the outcome assessments, pregnancy or lactating women, and history of adverse reaction(s) to local anesthetic or steroids
Marsh 2014	failed conservative treatment for 6 months male or female of skeletal maturity, age greater than 18 years lumbar spinal levels from L2 to S1 spinal stenosis at one or two consecutive levels no sign of segmental instability	spinal stenosis at more than two levels significantly compromised vertebral bodies at affected levels, e.g., previous surgery back or leg pain of unknown aetiology systemic or local infections severe obesity (BMI greater than 40) significant metabolic, autoimmune, peripheral vascular disease
Mobbs 2014	1) symptomatic LSS with radiculopathy (defined as well-localized lower-limb pain, weakness, or numbness), neurogenic claudication (defined as poorly localized back or lower-limb heaviness or numbness, with reduced tolerance for standing or ambulation), or urinary dysfunction 2) radiologically confirmed LSS (confirmed by either MRI or CT myelogram), caused by degenerative changes (facet joint hypertrophy, ligamentum flavum hypertrophy, and/or broadbased disc bulge) 3) canal stenosis at a maximum of 2 levels (that is, 1- or 2-level canal stenosis only)	1) were to undergo a concomitant fusion or instrumentation placement 2) had had previous lumbar surgeries at the same level 3) were to undergo lumbar laminectomy involving discectomy 4) had spondylolisthesis of any grade or degenerative scoliosis 5) had evidence of instability on dynamic radiographs

Moojen 2013	patients aged between 40 and 85 years with at least three months of intermittent neurogenic claudication due to single or two level degenerative lumbar canal stenosis and an indication for surgery were eligible. all patients were diagnosed as having intermittent neurogenic claudication by a neurologist in one of the participating hospitals. if magnetic resonance imaging showed a lumbar spinal canal stenosis, the consulting neurosurgeon could include patients as surgical candidates for the study.	patients with a cauda equina syndrome, a herniated disc needing discectomy, history of lumbar surgery, or significant scoliosis (Cobb angle >25°) or other spinal deformities
Musacchio 2016	1. radiographical confirmation of at least moderate lumbar stenosis, which narrows the central spinal canal at 1 or 2 contiguous levels from L1–L5 that require surgical decompression. Moderate stenosis is defined as more than 25% reduction of the anteroposterior dimension compared with the next adjacent normal level, with nerve root crowding compared with the normal level, as determined by the investigator on CT Scan or MRI. The patient may have, but is not required to have for inclusion in the study: a. facet hypertrophy and subarticular recess stenosis at the affected level(s); b. foraminal stenosis at the affected level(s); c. up to grade I stable degenerative spondylolisthesis (Meyerding classification) or equivalent retrolisthesis as determined by flexion/extension radiograph: i. for single-level disease, there may be up to a grade I stable spondylolisthesis or equivalent retrolisthesis at the affected level as determined on flexion/extension films by the investigator. ii. for 2-level disease, there may be up to a grade I stable spondylolisthesis or equivalent retrolisthesis at only 1 of the 2 contiguous affected levels, as determined on flexion/extension films by the investigator. Patients with up to grade I stable spondylolisthesis at 2 contiguous levels are excluded, but patients with up to grade I stable spondylolisthesis at 1 level and equivalent retrolisthesis at the adjacent level may be included.	• more than 2 vertebral levels requiring surgical decompression. • prior surgical procedure that resulted in translatory instability of the lumbar spine [as defined by White & Panjabi]. • more than 1 surgical procedure at any combination of lumbar levels. • prior fusion, implantation of a total disc replacement, complete laminectomy, or implantation of an interspinous process device at any lumbar level. • radiographically compromised vertebral bodies at any lumbar level(s) caused by current or past trauma or tumor (*e.g.* , compression fracture). • severe facet hypertrophy that requires extensive bone removal that would cause instability. • isthmic spondylolisthesis or spondylolysis (pars fracture). • degenerative lumbar scoliosis (Cobb angle > 25°). • disc herniation at any lumbar level requiring surgical intervention. • *Osteopenia*: A screening questionnaire for osteopenia, SCORE (simple calculated osteoporosis risk estimation), will be used to screen patients who require a DEXA bone mineral density measurement. If DEXA is required, exclusion will be defined as a DEXA bone density measured T score of ≤ -1.0 (The World Health Organization definition of osteopenia). • back or leg pain of unknown etiology. • axial back pain only, with no leg, buttock, or groin pain. • morbid obesity defined as a body mass index > 40. • pregnant or interested in becoming pregnant in the next 3 years. • known allergy to titanium, titanium alloys, or MR contrast agents. • active or chronic infection—systemic or local.

	d. mild lumbar scoliosis (Cobb angle up to 25º). 2. radiographical confirmation of the absence of angular or translatory instability of the spine at index or adjacent levels (instability as defined by White & Panjabi: Sagittal plane translation > 4.5 mm or 15% or sagittal plane rotation >15° at L1–L2, L2–L3, and L3—L4; >20° at L4–L5 based on standing flexion/extension radiographs). 3. VAS back pain score of at least 50 mm on a 100 mm scale. 4. neurogenic claudication as defined by leg/buttocks or groin pain that can be relieved by flexion such as sitting in a chair. 5. patient has undergone at least one epidural injection at any prior time point, and at least 6 mo of prior conservative care without adequate and sustained symptom relief. 6. age between 40 and 80 yr. 7. Oswestry Low Back Pain Disability Questionnaire score of at least 20/50 (40%). 8. appropriate candidate for treatment using posterior surgical approach. 9. psychosocially, mentally, and physically able to comply fully with this protocol, including adhering to scheduled visits, treatment plan, completing forms, and other study procedures. 10. personally signed and dated informed consent document prior to any study-related procedures indicating that the patient has been informed of all pertinent aspects of the trial.	• chronically taking medications or any drug known to potentially interfere with bone/soft tissue healing (*e.g.*, steroids), not including a Medrol (Methylprednisolon) dose pack. • history of significant peripheral neuropathy. • significant peripheral vascular disease (*e.g.*, with diminished dorsalis pedis or posterior tibial pulses). • unremitting back pain in *any* position. • uncontrolled diabetes. • known history of Paget disease, osteomalacia, or any other metabolic bone disease (excluding osteopenia, which is addressed earlier). • cauda equina syndrome, defined as neural compression causing neurogenic bowel (rectal incontinence) or bladder (bladder retention or incontinence) dysfunction. • fixed and complete motor, sensory, or reflex deficit. • rheumatoid arthritis or other autoimmune diseases. • known or documented history of communicable disease, including AIDS, HIV, active hepatitis. • active malignancy: a patient with a history of any invasive malignancy (except nonmelanoma skin cancer), unless he/she has been treated with curative intent and there has been no clinical signs or symptoms of the malignancy for at least 5 years. Patients with a primary bony tumor are excluded as well. • prisoner or ward of the state. • subject has a history of substance abuse (*e.g.*, recreational drugs, narcotics, or alcohol). • subject is currently involved in a study of another investigational product for similar purpose. • currently seeking or receiving workman's compensation. • in active spinal litigation.
Nam 2011	1) cases with pain that increased with lumbar extension and decreased with lumbar flexion 2) patients with radiating pain present at least below the knee joint 3) cases with a thoracolumbar scoliosis greater than 10 degrees, visible on the standard Rx in the standing anterior-posterior (AP) and lateral views of the whole spine, including the hip joint and the cervical spine, or in the standing lateral bending views of the lumbar spine. the subjects were patients who were found to have spinal stenosis on both CT and MRI examinations of the lumbar spine performed for the nerve-root location in those cases.	1) patients with any systemic inflammatory disease or diabetes 2) patients taking anticoagulant medication 3) patients who had previously experienced side effects from the use of lidocaine or contrast dye 4) patients with any known or suspected infectious disease 5) patients who found it difficult to regularly visit the hospital because of general bad health 6) patients with a skin disorder rending them unsuitable for injection administration on the injection site 7) cases with a mental health problem who were unable to complete a questionnaire 8) patients who had received a steroid injection in the three months prior to the beginning of the study 9) cases with degenerative spondylolisthesis, osteoporosis or compression fracture

	degenerative lumbar scoliosis patients were included in the study only when their radicular pain resulted from lateral foraminal stenosis.	10) patients who had received surgical treatment of the thoracolumbar region or cases with cancer metastasis to the thoracolumbar site or with spinal deformity caused by metabolic disease in contrast, patients with spinal stenosis or neurogenic claudicaiton were excluded from the study. we also excluded cases with neurological symptoms such as cauda equine syndrome, and patients that needed nonsteroidal anti-inflammatory drugs other than acetaminphen or low dose aspirin, as well as those who continued to receive other conservative treatments such as physiotherapy over the clinical trial period.
Park 2019	participants' age between 30 and 80 years degenerative lumbar stenosis with radiating pain to lower extremities (score of visual analog scale >4) definite lumbar central stenosis (Schizas grade ≥B) on magnetic resonance imaging	spondylolisthesis (≥Meyer grade II) history of lumbar spinal surgery for spinal stenosis or instability at the same level stenosis caused by a herniated intervertebral disc degenerative lumbar scoliosis (Cobb angle >20°) other spinal diseases (e.g., ankylosing spondylitis, spine tumor, fracture, or neurologic disorders) psychologic disorders (e.g., dementia, intellectual disability, or drug abuse) other disorders that the surgeon considered to make participation inappropriate
Rajasekaran 2013	degenerative LCS affecting 3 or less levels, typical neurogenic claudication symptoms, magnetic resonance image demonstrating good clinical correlation, and failure of conservative methods of treatment for a minimum period of 6 months.	spondylolisthesis with slip grade 2 or greater (Meyerding grade), instability at the level of stenosis (as defined by > 3-mm translation or > 10 ° angular change on flexion extension lateral radiographs), concomitant symptomatic cervical or thoracic stenosis, and comorbidities such as cardiopulmonary insufficiency, peripheral neuropathy, peripheral vascular disease, prior lumbar spine surgery, and severe hip or knee disease.
Ruetten 2009	the following clinical inclusion criteria applied: neurogenic claudication with unilateral leg pain with or without paresis; back pain with maximum score of 20 of 100 points on the VAS; and conservative therapy exhausted or no longer indicated due to the symptoms the imaging inclusion criteria were as follows: monosegmental recess stenosis; no foraminal stenosis in the lower level; no disc herniation; degenerative spondylolisthesis with maximum Meyerding Grade I; no multidirectional	

	rotation slide; scoliosis, maximum curvature 20°; and no prior surgery in the same segment	
Schmidt 2018	age >40 yrs radiographic confirmation of clinical symptoms of at least moderate degenerative spinal stenosis, w/ constriction of the central spinal canal of 1 or 2 adjacent segments in the L3–5 region w/ the need for decompression. Diagnosis must include: 1. minimum of 3 mos of conservative therapy w/out improvement of symptoms 2. radiographic confirmation of no translational instability in main segment as well as in adjacent segments (dynamic translational instability ≤3 mm) 3. VAS back pain score ≥50 mm (out of 100) 4. ODI score of ≥18 (out of 45; 40%) if necessary, additional decompression in the adjacent segment(s) may be performed, avoiding any instability in the affected segment. in addition, the following may exist but are not required: i. hypertrophy of the facet joints & subarticular recessus stenosis in the relevant segment ii. stenosis of the foramen in the relevant segment iii. stable retrolisthesis up to grade I verified by flexion-extension radiographic films mental & physical ability of patient to follow the protocol (i.e., compliance w/ time schedule & treatment plan, able to fill in CRF pages & to undergo further study procedures)	1. preceding fusion or decompression surgery of the lumbar spine or preceding nucleotomy of the segments of concern (also if nucleotomy becomes necessary during surgery) 2. radiographically confirmed damage of the vertebral body in the segment of concern in the lumbar spine (e.g., osteoporotic compression fracture or because of tumors) 3. isthmic & degenerative spondylolisthesis (anterolisthesis; retrolisthesis > grade I) or spondylolysis (pars fracture) 4. degenerative lumbar scoliosis (>25°) 5. adipositas (obesity); defined as a BMI >40 6. pregnancy, or wish to get pregnant during the course of the study 7. known allergy to titanium & titanium alloys 8. florid infections—both systemic & local 9. history of severe peripheral neuropathy 10. significant peripheral vascular disease (claudicatio intermittens ≥ stage 2b) 11. Paget disease or osteomalacia or other metabolic bone disorders 12. cauda equina syndrome 13. communicable diseases, including HIV, active hepatitis 14. patients who are lawfully kept in an institution 15. patients who, in the opinion of the investigator, will be inappropriate for inclusion in this clinical trial or who will not comply w/ requirements of the study 16. patients who participated in a clinical observation or therapy w/ radiography during the last 10 yrs 17. patients who participate(d) in another clinical trial (w/in the last 4 wks) that might influence the safety & effectiveness assessment of this trial
Skoro 2016	degenerative lumbar disease of ≥2 levels causing neurogenic claudication with unilateral or bilateral radiculopathy, shortened walking distance of <100 m, and the inability to stand still for >5 minutes; MRI confirmation of absolute lumbar spinal stenosis measured as the surface of the dural sac at the most compressed level ≤75 mm^2 in at least 1 level; and symptom	scoliosis >20°; anterolisthesis >16%; retrolisthesis >12%; sagittal kyphosis >12°; previous surgery of the lumbar spine; lumbar spine trauma; and the presence of lumbar spinal tumors, infections, and cauda equina syndrome

	duration for a minimum of 6 months with no improvement with conservative therapy	
Song 2016	patients diagnosed with spinal stenosis underwent MRI and electrodiagnostic examinations. Included patients had an anterior-posterior (AP) diameter of the spinal canal of less than 12 mm confirmed through sagittal imaging, and an AP foraminal diameter of less than 3 mm confirmed through parasagittal imaging, both by MRI, and were found to have abnormal somatosensory evoked potentials (SEPs)	we excluded patients who had a lower limb vascular disorder, a psychological problem, another musculoskeletal disorder or symptoms of a neurogenic bladder or bowel.
Strömqvist 2013	the patients had to have MRI verified spinal stenosis on 1 or 2 levels in the lumbar spine symptoms of neurogenic claudication for minimum 6 months elicited by walking and relieved by flexion of the spine or sitting down patient age 40 years or more spinal stenosis was allowed to be present at maximum 2 levels and minor spondylolisthesis (Meyerding, grade 1)	the L5–S1-level excluded previous spine surgery (except for successful disc surgery), infection or malignant disorder, and osteoporosis diagnosed before referral for surgery and subjected to medical treatment
Thom 2005	1) symptoms of neurogenic claudication or radiculopathy 2) radiological/neuroimaging evidence of degenerative lumbar stenosis 3) absence of associated pathological entities such as disc herniations or instability 4) no history of surgery for lumbar stenosis or lumbar fusion symptoms were considered refractory to nonsurgical management if conservative measures, particularly nonsteroidal antiinflammatory drug and physical therapies, had been administered for at least 3 months without sufficient improvement patients presenting with stable spondylolisthesis or a history of surgery for herniated lumbar discs were not excluded	we excluded from outcome analysis three patients who required discectomies due to significant intraoperatively noted discogenic nerve compression, which had not been identified on preoperative imaging studies
Watanabe 2011	1) presence of neurogenic claudication; 2) symptoms persistent for more than 6 months despite conservative therapy;	1) spinal canal stenosis due to congenital, spondylolytic, traumatic, and iatrogenic causes; 2) any previous operation in the lumbar area;

	3) clinical symptoms and neurological signs in the lower limbs corresponding to the level of stenosis on MR imaging or myelography; 4) 1- or 2-level decompression necessary. radiographic instability of the lumbar spine and degenerative spondylolisthesis were not regarded as exclusion criteria	3) presence of other specific spinal disorders (such as ankylosing spondylitis, neoplasm, or metabolic diseases); 4) intermittent claudication resulting from peripheral arterial disease; 5) severe osteoarthrosis or arthritis in the lower limbs; 6) neurological disease causing impaired lower-limb function, including diabetic neuropathy; 7) psychiatric disorders; 8) multilevel spinal canal stenosis requiring decompression at 3 or more levels
Weinstein 2007	all patients had neurogenic claudication or radicular leg pain with associated neurologic signs, spinal stenosis shown on cross-sectional imaging, and degenerative spondylolisthesis shown on lateral radiographs obtained with the patient in a standing position. The patients had had persistent symptoms for at least 12 weeks and had been confirmed as surgical candidates by their physicians. patients with adjacent levels of stenosis were eligible	patients with spondylolysis and isthmic spondylolisthesis
Weinstein 2008	a history of neurogenic claudication or radicular leg symptoms for at least 12 weeks and confirmatory cross-sectional imaging showing lumbar spinal stenosis at one or more levels; all patients were judged to be surgical candidates	patients with lumbar instability (which was defined as translation of more than 4 mm or 10 degrees of angular motion between flexion and extension on upright lateral radiographs)
Yagi 2009	symptoms of neurogenic claudication referable to the lumbar spine failure of conservative treatments; minimum 3 mos absence of associated pathological condition; 1-level spondylosis the presence of Grade I spondylolisthesis without segmental instability was not considered a contraindication to this study	
Zucherman 2005	patients had to be at least 50 years old and have leg, buttock, or groin pain with or without back pain that was relieved during flexion. to identify a study population of patients with more moderate symptoms of NIC, patients had to be able to walk at least 50 feet.	patients could not have a fixed motor deficit, cauda-equina syndrome, previous lumbar surgery of the stenotic level, or spondylolisthesis greater than grade I on a scale of I to IV at the affected level(s).

Did not mention failure of conservative treatments

Amundson 2000, Delitto 2015, Forsth 2016, Friedly 2016, Ghogawala 2016, Grob 1995, Gurelik 2012, Ko 2019, Koc 2019, Liu 2013, Slatis 2011, Mobbs 2014, Moojen 2013, Park 2019, Song 2016, Stromqvist 2013

Without typical symptom.

The typical symptom of LSS is the neurogenic claucatio, which a pain progressing from the back of the buttock down along the legs (one leg in case of unilateral stenosis) during walking that improves while sitting and during a forward bending of the trunk

Amundsen 2000, Azzazi 2010, Cho 2007, Elsheikh 2016, Forsth 2016, Grob 1995, Koc 2009, Lin 2013, Slatis 2011, Manchikanti 2009, Manchikanti 2012, Manchikanti 2015, Marsh 2014, Nam 2011, Park 2019, Schmidt 2018, Song 2016, Zucherman 2005

Calendar year (After 2000)

Amundsen 2000, Grob 1995

Not mention instability or have instability

Amundsen 2000, Anderson 2006, Azzazi 2010, Benyamin 2016, Brown 2012, Delitto 2015, Elsheikh 2016, Friedly 2016, Hallett 2007, Karm 2018, Kim 2013, Koc 2009, Koh 2013, Komp 2015, Makoto 1998, Manchikanti 2009, Manchikanti 2012, Manchikanti 2015, Moojen 2015, Musacchio 2016, Nam 2011, Ruetten 2009, Skoro 2016, Slatis 2011, Song 2016, Strömqvist 2013, Watanabe 2011, Weinstein 2007, Zucherman 2005

Appendix S5: Certainty of evidence and definitions

• High certainty—We are very confident that the true effect lies close to that of the estimate of the effect

• Moderate certainty—We are moderately confident in the effect estimate. The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different

• Low certainty—Our confidence in the effect estimate is limited. The true effect may be substantially different from the estimate of the effect

• Very low certainty—We have very little confidence in the effect estimate. The true effect is likely to be substantially different from the estimate of effect

Appendix S6: Study, Patient and Intervention Characteristics

First Author, Year	Trial Registry Identifier	Funding Source Type	Single/Multicentre	Study Location (s)	#Study Centres	Study Period	Study Length (mos.)	Outcomes reported
Amundsen, 2000[1]	NR	NR	Single	Norway	1	Dec 1984 to Sep 1987	120	OP, PF, TW, AE, DE
Anderson, 2006[2]	NR	NR	Multi	USA	9	NR	24	PF, GIR, AE
Azzazi, 2010[3]	NR	NR	Single	Egypt	1	Mar 2005 to May 2007	24	LP, BP, PF, AE
Benyamin, 2016[4]	NCT02093520	Commercial	Multi	USA	26	Jun 2014 to Apr 2015	12	OP, PF, GIR, TW, AE, DE
Brown, 2012[5]	NCT00995371	Commercial	Single	USA	1	Sep 2009 to Jan 2011	6	OP, PF, GIR
Cavuolu, 2007[6]	NR	NR	NR	Turkey	NR	Jan 2000 to Jan 2002	Mean:65	OP, PF, AE, DE
Celik, 2010[7]	NR	NR	Single	Turkey	1	Jul 2001 to May 2003	60	LP, BP, PF, TW, MB, AE
Cho, 2007[8]	NR	NR	Single	China	1	NR	10 - 18	BP, PF, GIR, AE
Delitto, 2015[9]	NCT00022776	Government	Multi	USA	2	Nov 2000 to Oct 2005	24	PF, TW, AE, DE
Elsheikh, 2016[10]	NR	NR	Single	Egypt	1	Jan 2013 to Dec 2014	12	OP, PF, TW, MB, AE
Forsth, 2016[11]	NCT01994512	Government	Multi	Sweden	NR	Oct 2006 to Jun 2012	24	LP, BP, PF, GIR, TW, MB, AE, DE
Friedly, 2014[12,13]	NCT01238536	Government	Multi	USA	16	Apr 2011 to Jun 2013	12	LP, BP, PF, TW, AE
Ghogawala, 2016[14]	NCT00109213	Commercial	Multi	USA	5	Mar 2002 to Aug 2009	48	PF, TW, AE, DE
Grob, 1995[15]	NR	NR	Single	Switzerland	1	Nov 1989 to Nov 1990	Mean: 28	OP, MB, AE
Gurelik, 2012[16]	NR	NR	Single	Turkey	1	Jan 2006 to Feb 2009	Mean: 9.1	PF, MB, AE
Haddadi, 2016[17]	NR	NR	NR	Iran	NR	NR	12	LP, BP, PF, GIR, AE, DE
Hallett, 2007[18]	NR	NR	NR	UK	NR	Jan 1998 to Aug 2011	60	BP, PF, GIR, TW, AE, DE

Kang, 2019[19]	NR	Government	Single	Korea	1	Jan 2015 to Dec 2016	6	BP, PF, GIR, AE
Karm, 2018[20]	KCT 0002093	NR	Single	Korea	1	Jan 2014 to Jun 2016	6	LP, BP, PF, GIR, TW, AE
Kim, 2013[21]	NR	NR	Single	Korea	1	Jul 2010 to Aug 2011	3	LP, BP, PF, MB, TW, AE
Ko, 2019[22]	NR	NR	NR	Korea	NR	Jan 2015 to Jun 2016	24	LP, BP, PF, DE
Koc, 2009[23]	NR	NR	NR	Turkey	NR	NR	6	OP, PF
Koh, 2013[24]	KCT0000500	NR	Single	Korea	1	Jan 2011 to Jan 2012	6	LP, PF, GIR, TW, AE
Komp, 2015[25]	NR	NR	NR	Germany	NR	NR	24	LP, BP, PF, AE, DE
Liu, 2013[26]	NR	NR	NR	China	NR	NR	24	LP, BP, PF
Lønne, 2015[27]	NCT00546949	Government	Multiple	Norway	6	Jun 2007 to Sep 2011	24	LP, BP, PF, GIR, TW, AE, DE
Makoto, 1998[28]	NR	NR	NR	Japan	NR	NR	3	MB
Manchikanti, 2009[29]	NCT00370994	Commercial	Single	USA	1	Start from Jan 2006	12	OP, PF, AE, DE
Manchikanti, 2012[30-32]	NCT00370799	Hospital	Single	USA	1	Jan 2007 to Dec 2009	24	OP, PF, TW
Manchikanti, 2015[33,34]	NCT00681447	Commercial	Single	USA	1	Jan 2008 to Jul 2011	24	OP, PF, TW, AE, DE
Marsh, 2014[35]	NR	NR	Single	UK	1	NR	48	BP, PF, DE
Mobbs, 2014[36]	NR	NR	Single	Australia	1	2007 to 2009	Mean: 40.6	LP, PF, GIR, TW, AE
Moojen, 2015[37,38]	NTR1307	Commercial	Multiple	Netherland	5	Oct 2008 to Sep 2011	24	LP, BP, PF, MB, TW, AE
Musacchio, 2016[39-41]	NR	Commercial	Multiple	USA	21	Oct 2006 to Mar 2010	60	LP, BP, PF, TW, AE
Nam, 2011[42]	NR	Government	Single	Korea	1	Jan 2009 to Jun 2010	3	OP, PF, TW
Park, 2019[43]	NCT03302507	Hospital	Single	Korea	1	Nov 2017 to Aug 2018	0.5	TW, AE
Rajasekaran, 2013[44]	NR	Government	Single	India	1	NR	Mean: 14.2	LP, BP, PF, AE

Ruetten, 2009[45]	NR	NR	NR	Germany	NR	2003 to 2005	24	LP, BP, PF, GIR, TW, AE, DE
Schmidt, 2018[46]	NCT01316211	Commercial	Multiple	Germany	7	Mar 2008 to Jul 2014	24	LP, BP, PF, AE, DE
Skoro, 2016[47]	NR	NR	NR	Croatia	NR	Dec 2000 to Mar 2005	96	PF, AE
Slatis, 2011[48,49]	NR	Hospital	Multiple	Finland	4	Dec 1997 to Mar 2001	72	LP, BP, PF, MB, TW, AE, DE
Song, 2016[50]	NR	NR	NR	Korea	NR	Oct 2012 to Jan 2014	3	OP, PF
Strömqvist, 2013[51]	NR	NR	Multiple	Sweden	3	NR	24	LP, BP, PF, GIR, TW, AE, DE
Thom, 2005[52]	NR	NR	NR	Germany	NR	NR	Mean: 15.5	OP, LP, BP, PF, GIR, MB, TW, AE
Watanabe, 2011[53]	NR	NR	Single	Japan	1	Dec 2004 to Dec 2005	0.25	PF, TW, AE
Weinstein, 2007[54,55]	NCT00000409	Government	Multiple	USA	13	Mar 2000 to Mar 2005	24	QoL, PF, TW, GIR, AE, DE
Weinstein, 2008[56]	NCT00000411	Government	Multiple	USA	13	Mar 2000 to Mar 2005	24	PF, GIR, TW, AE, DE
Yagi, 2009[57]	NR	NR	NR	Japan	NR	NR	12	PF, LP, AE
Zucherman, 2005[58-60]	NR	Commercial	Multiple	USA	9	May 2000 to Jul 2001	24	PF, GIR, TW, AE, DE

First author, Year	Level	DS	% Instability	Type	Study N	% Male	Patient age	% Smoker	BMI	Trial arms
Amundsen, 2000	1 or more	NR	NR	C/L/F	31	NR	NR	NR	NR	CD vs Cons
Anderson, 2006	1 or 2	LG	NR	NR	75	40	Mean: 70.1	NR	Mean: 27.4	ID vs EpiS
Azzazi, 2010	1 or 2	LG	NR	C/L	60	31.7	Mean: 56.7	43.3	Mean: 28	CD+Fu vs ID
Benyamin, 2016	1 or 2 or 3	LG	NR	C/L/F	302	43.7	Mean: 75.3	NR	NR	Endo vs EpiS
Brown, 2012	1 or more	LG	NR	C	38	55.3	Mean: 76.2	NR	NR	Endo vs EpiS
Cavuolu, 2007	2 or 3 or 4	LG	0[a]	C	100	39	Mean: 69.2	NR	NR	Two similar MPD
Celik, 2010	1 or more	LG	0[b]	NR	71	53.5	Mean: 60	NR	NR	MPD vs CD
Cho, 2007	1 or more	LG	0[c]	NR	70	44.3	Mean: 60.2	NR	NR	MSD vs CD
Delitto, 2015	NR	LG	NR	C	169	52.1	Mean: 68.2	7.7	Mean: 31.3	CD vs Cons
Elsheikh, 2016	1 or more	LG	NR	C	132	65.2	Mean: 57	NR	NR	EpiAS vs EpiASC
Forsth 2016	1 or 2	LG+HG	0[c]	C	233	33.5	Mean: 66.9	15	NR	CD+Fu vs CD
Friedly 2016	NR	LG	NR	C	400	44.8	Mean: 68	14.3	Mean: 30.4	EpiA vs EpiAS
Ghogawala, 2016	NR	LG	0[d]	C	66	19.7	Mean: 66.6	NR	NR	CD vs CD+Fu
Grob 1995	NR	NR	0[e]	NR	45	46.7	NR	NR	NR	CD vs CD+Fu
Gurelik, 2012	NR	NR	0[f]	C	52	40.4	Mean: 59.1	NR	NR	MPD vs CD
Haddadi, 2016	NR	LG	0[a]	C	120	54.2	Mean: 67.7	NR	Mean: 25.3	MPD vs MSD vs CD
Hallett, 2007	1	LG	NR	F	44	54.5	Mean: 57	27.3	NR	CD vs CD+Fu vs CD+Fu

Kang, 2019	1	NR	0[c]	C	62	51.6	Mean: 66.2	NR	NR	Endo vs MPD
Karm, 2018	1 or 2 or 3	NR	NR	C	44	59.1	Mean: 65.8	NR	Mean: 24.3	BT vs EpiAS
Kim, 2013	NR	NR	NR	C	62	54.8	Mean: 64.9	NR	Mean: 24	BT vs EpiAS
Ko, 2019	NR	NR	0[g]	C	50	36	Mean: 67.2	NR	Mean: 24	CD vs MPD
Koc 2009	NR	NR	NR	NR	29	72.4	Mean: 58.9	NR	NR	Cons vs EpiAS vs Cons
Koh 2013	1	LG	NR	L	53	28.3	Mean: 64.9	NR	Mean: 25.1	EpiS vs EpiSH
Komp 2015	NR	LG	NR	C	160	43.1	Mean: 62	NR	NR	Endo vs MPD
Liu 2013	1 or 2	LG	0[c]	C	56	58.9	Mean: 60.3	NR	NR	MSD vs MPD
Lønne 2015	1 or 2	LG	0[c]	C	81	49.4	Mean: 67	25.9	Mean: 28	ID vs MPD
Makoto 1998	NR	NR	NR	C/L	53	71.7	Mean: 70.4	NR	NR	Epi vs EpiA vs EpiAS
Manchikanti, 2009	NR	NR	NR	NR	50	42	Mean: 61.5	NR	Mean: 29.7	EpiAS vs EpiASH
Manchikanti, 2012	1 or more	NR	NR	C	100	41	Mean: 56.4	NR	Mean: 30.4	EpiA vs EpiAS
Manchikanti, 2015	1 or more	NR	NR	C	120	43.3	Mean: 52.3	NR	Mean: 30.4	EpiA vs EpiAS
Marsh, 2014	NR	LG	0[c]	NR	60	50	Mean: 58	NR	NR	CD+ID vs CD
Mobbs, 2014	1 or 2	LG	0[c]	C/L	54	31.5	Mean: 69.3	22.2	NR	MPD vs CD
Moojen, 2015	1 or 2	NR	NR	C	159	54.1	Mean: 65	NR	Mean: 27.5	ID vs CD
Musacchio, 2016	1 or 2	LG	NR	C	322	49.1	Mean: 62.8	NR	Mean: 29.7	CD+ID vs CD+Fu
Nam, 2011	NR	LG	NR	NR	36	25	Mean: 73.3	NR	Mean: 23.1	EpiAS vs EpiA
Park, 2019	NR	LG	0[c]	C	64	48.4	Mean: 66.7	28.1	Mean: 25.1	Endo vs MPD

Rajasekaran, 2013	1 or 2 or 3	LG	0[h]	C	51	58.8	Mean: 56	NR	NR	CD vs MSD
Ruetten, 2009	NR	LG	NR	C	192	54.2	Mean: 64	NR	NR	Endo vs MPD
Schmidt, 2018	1 or 2	LG	0[i]	C	225	46.2	Mean: 68	NR	Mean: 29.2	CD+ID vs CD
Skoro, 2016	2 or more	LG	NR	C	44	63.6	Mean: 64.8	NR	NR	MPD vs CD+Fu
Slatis, 2011	1 or more	NR	NR	NR	94	33	Mean: 62.5	NR	Mean: 27.5	CD vs Cons
Song, 2016	NR	NR	NR	C/L	29	48.3	Mean: 60	NR	NR	EpiS vs EpiAS
Strömqvist, 2013	1 or 2	LG	NR	C	100	56	Mean: 69	NR	NR	ID vs CD
Thom, 2005	NR	LG	0[a]	C	120	44.2	Mean: 68	NR	Mean: 28.7	MPD vs MPD vs CD
Watanabe, 2011	1 or 2D	NR	NR	C	34	52.9	Mean: 69.9	NR	NR	MSD vs CD
Weinstein, 2007	1 or more	HG	9[g]	C/L/F	301	34	Mean: 66.0	8	Mean: 29.1	CD+Fu vs Cons
Weinstein, 2008	1 or more	LG	0[g]	C/L/F	278	62	Mean: 65.5	12	Mean: 29.8	CD vs Cons
Yagi, 2009	1	LG	0[c]	NR	41	34.1	Mean: 72	NR	NR	MPD vs CD
Zucherman, 2005	1 or 2	LG	NR	NR	191	57.1	Mean:69.6	NR	Mean: 28.1	ID vs EpiS

a: spinal instability was defined as sagittal-plane translation of 5 mm or more documented on flexion–extension radiography.

b: lumbar instability measured on flexion and extension lateral radiographs is defined by greater than 4mm of translation (8%) or greater than 10 to 12 degrees of angular displacement.

c: no definition.

d: motion of >3 mm at the level of listhesis, as measured on flexion-extension radiographs of the lumbar spine.

e: (1) a concomitant slip of a vertebra of more than five millimeters or another gross deformity such as rotational instability characterized by more than five millimeters of lateral offset on the anteroposterior roentgenogram or degenerative scoliosis. (2) spondylolysis with an osseous defect of the pars interarticularis, or (3) a previous operation on the lumbar spine.

f: 1) anterior translation greater than 8% (L1-2 to L4-5) or greater than 6% (L5-S1) of the vertebral body width; 2) posterior translation greater than 9% (L1-S1); 3) angular displacement (sagittal rotation) in flexion greater than - 9° (L5-S1) or greater than 1° (L5-S1).

g: Patients confirmed having instability when > 4-mm translation or > 10° angulation was viewed in lateral flexion and extension images.

h: by > 3-mm translation or > 10 ° angular change on flexion extension lateral radiographs.

i: radiographic confirmation of translational instability in the main segment as well as in adjacent segments (dynamic translational instability ≤ 3 mm).

Appendix S7: Descriptive summary of studies not included in analysis

First author, Year	Trial Arms	Summary of findings
Cavuolu, 2007	Two similar MPD	Analysis of clinical outcome showed no statistical differences between two groups
Hallett, 2007	CD vs CD+Fu vs CD+Fu	no significant additional benefit was found with the more complex surgery
Makoto 1998	Epi vs EpiA vs EpiAS	Epidural steroid injection has no beneficial effect on claudication associated with spinal canal stenosis as compared with epidural block with a local anesthetic alone
Koh, 2013	EpiS vs EpiSH	Superior short-term pain relieving efficacy, but limited long-term effects of hypertonic saline, when added to transforaminal epidural steroid injections

Appendix S8: Results of risk of bias assessment (n= 45 randomised controlled trials)

First author, Year	Domain 1	Domain 2	Domain 3	Domain 4	Domain 5	Overall risk of bias
Amundsen 2000	Some concerns	Some concerns	Low	High	Some concerns	High
Anderson 2006	Low	Some concerns	High	High	Some concerns	High
Azzazi 2010	Some concerns	Some concerns	Low	Low	Some concerns	Some concerns
Benyamin 2016	Some concerns	Some concerns	Low	High	Low	High
Brown 2012	Some concerns	Low	Low	Low	Low	Some concerns
Celik 2010	Some concerns	Some concerns	Low	Low	Some concerns	Some concerns
Cho 2007	Some concerns	High	High	High	Some concerns	High
Delitto 2015	Low	Low	Low	Low	Low	Low
Elsheikh 2016	Low	Low	Low	Low	Some concerns	Some concerns
Forsth 2016	Some concerns	Some concerns	Low	Some concerns	Low	Some concerns
Friedly 2016	Low	Low	Low	Low	Low	Low
Ghogawala 2016	Low	Low	High	Low	Low	High
Grob 1995	Some concerns	Some concerns	Low	High	Some concerns	High
Gurelik 2012	Some concerns	Some concerns	Low	Low	Low	Some concerns
Haddadi 2016	Some concerns	Some concerns	Low	High	Some concerns	High
Kang 2019	Some concerns	Low	Low	Low	Some concerns	Some concerns
Karm 2018	Some concerns	Low	Low	Low	Low	Some concerns
Kim 2013	Low	Low	Low	Low	Low	Low
Ko 2019	Some concerns	Some concerns	Low	Some concerns	Some concerns	Some concerns
Koc 2009	Some concerns	Some concerns	Low	Low	Some concerns	Some concerns
Komp 2015	Low	Some concerns	Low	Low	Some concerns	Some concerns
Liu 2013	Some concerns	Some concerns	Low	High	Some concerns	High
Lonne 2015	Some concerns	Some concerns	Low	High	Low	High
Slatis 2011	Low	Low	Low	High	Low	High
Manchikanti 2009	Low	Low	Low	Low	Low	Low
Manchikanti 2012	Low	Low	Low	Low	Low	Low
Manchikanti 2015	Low	Low	Low	Low	Low	Low
Marsh 2014	Some concerns	Some concerns	Low	High	Some concerns	High
Mobbs 2014	Some concerns	Some concerns	High	Low	Low	High
Moojen 2013	Low	Low	Low	Low	Low	Low

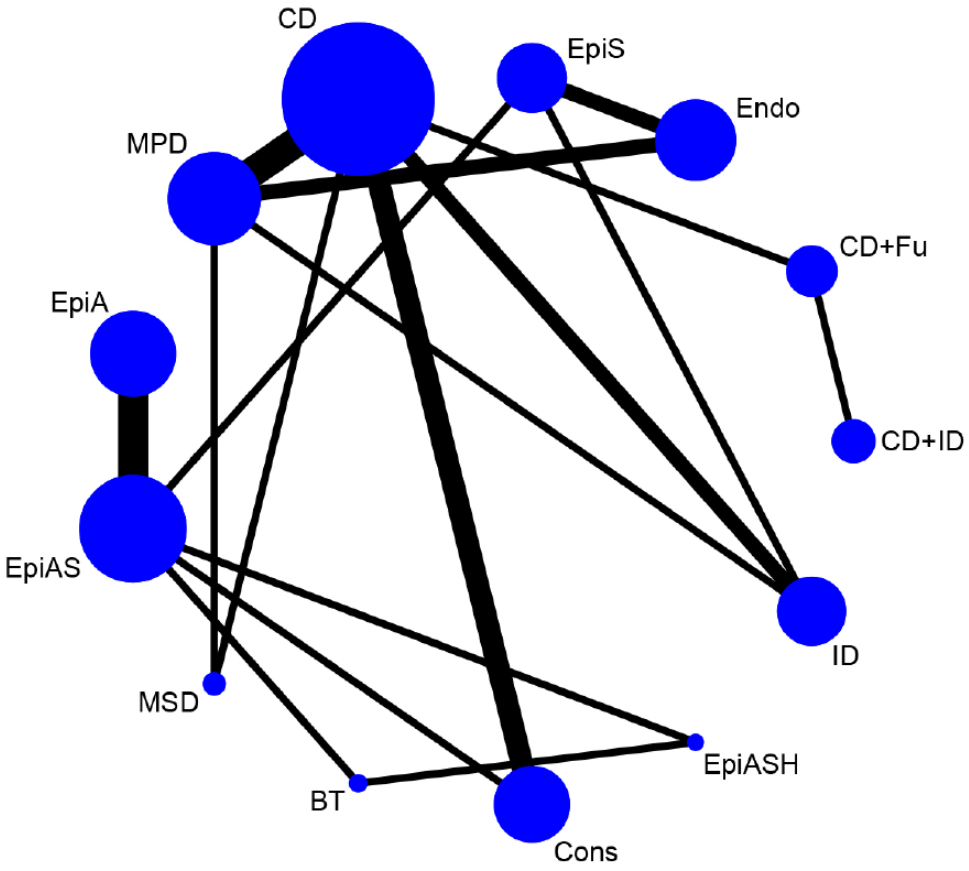
Musacchio 2016	Low	Low	Low	Low	Some concerns	Some concerns
Nam 2011	Some concerns	Some concerns	High	High	Some concerns	High
Park 2019	Low	Some concerns	Low	Low	Low	Some concerns
Rajasekaran 2013	Some concerns	Some concerns	Low	Low	Some concerns	Some concerns
Ruetten 2009	Some concerns	Some concerns	Low	Low	Some concerns	Some concerns
Schmidt 2018	Low	Some concerns	Low	High	Low	High
Skoro 2016	Some concerns	High	High	High	Some concerns	High
Song 2016	Some concerns	Some concerns	Low	High	Some concerns	High
Strömqvist 2013	Low	Some concerns	Low	High	Some concerns	High
Thom 2005	Low	Some concerns	Low	High	Some concerns	High
Watanabe 2011	Low	Some concerns	Low	High	Some concerns	High
Weinstein 2007	Low	Some concerns	Low	Low	Low	Some concerns
Weinstein 2008	Low	Some concerns	Low	Low	Low	Some concerns
Yagi 2009	Some concerns	Some concerns	Low	High	Some concerns	High
Zucherman 2005	Low	High	Low	High	Some concerns	High

Domain 1: Risk of bias arising from the randomization process; Domain 2: Risk of bias due to deviations from the intended interventions (effect of assignment to intervention); Domain 3: Missing outcome data; Domain 4: Risk of bias in measurement of the outcome; Domain 5: Risk of bias in selection of the reported result.

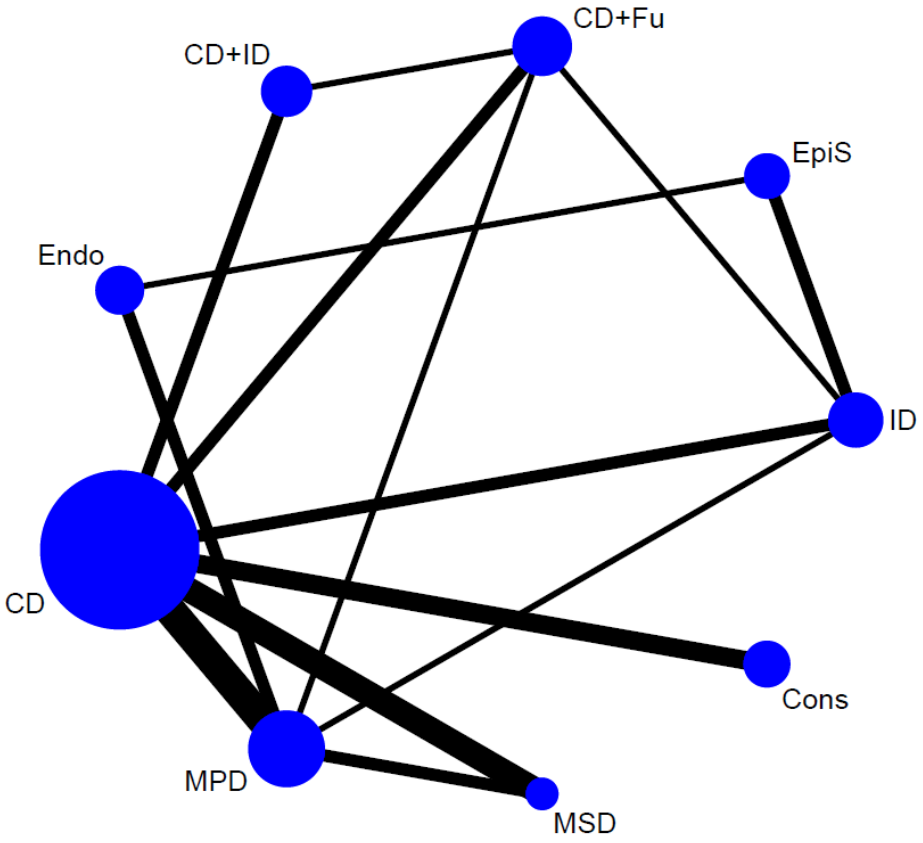
Appendix S9: Network plots

For each outcome, the architecture of the treatment comparisons equals to the geometry of the network, which is assessed through the network plot. The network plot contains two elements: node and line. Each node represents an intervention, and each line represents a direct comparison. The size of one node represents the number of patients in that intervention and the thickness of one line represents the number of studies in that comparison.

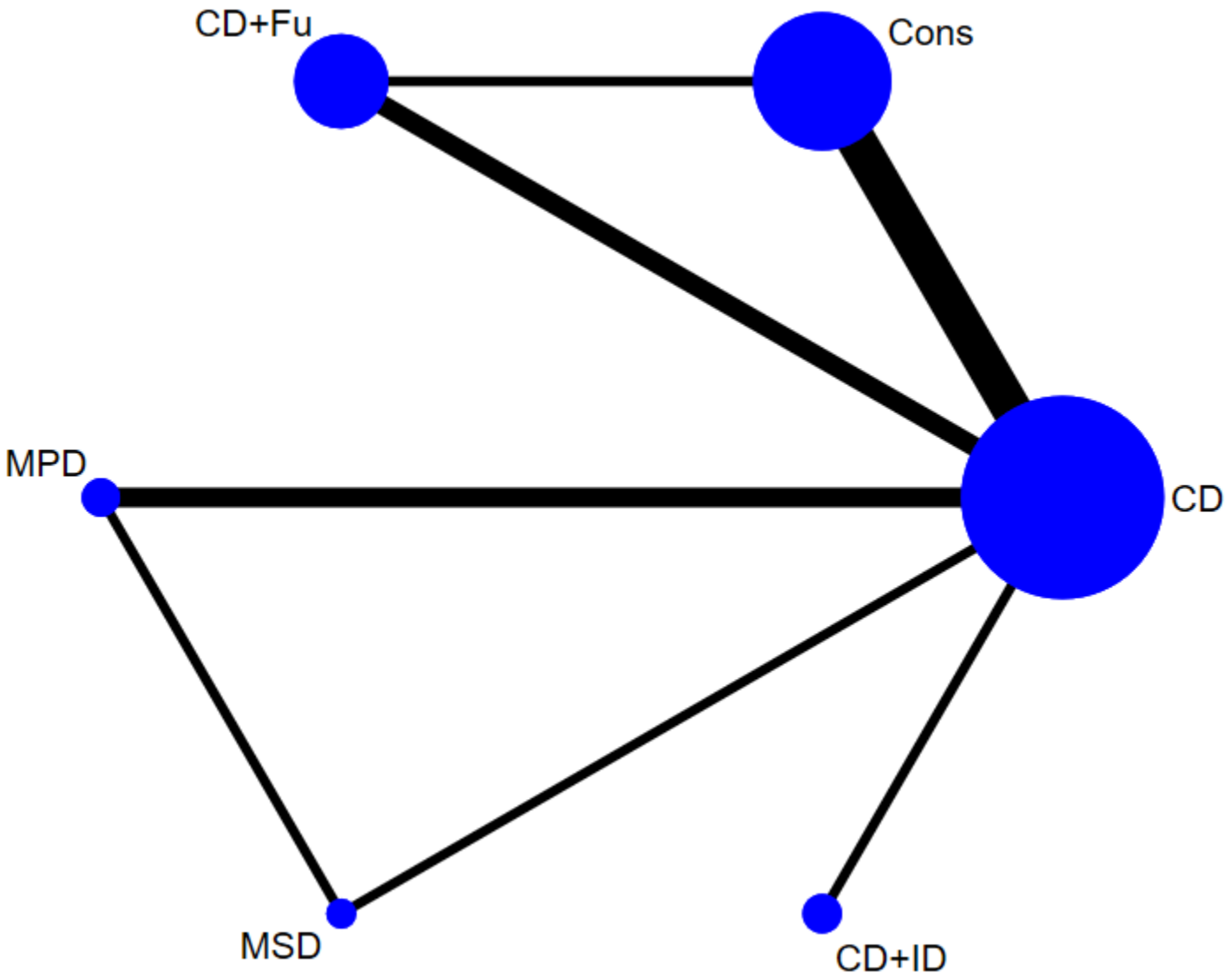
Short-term physical function



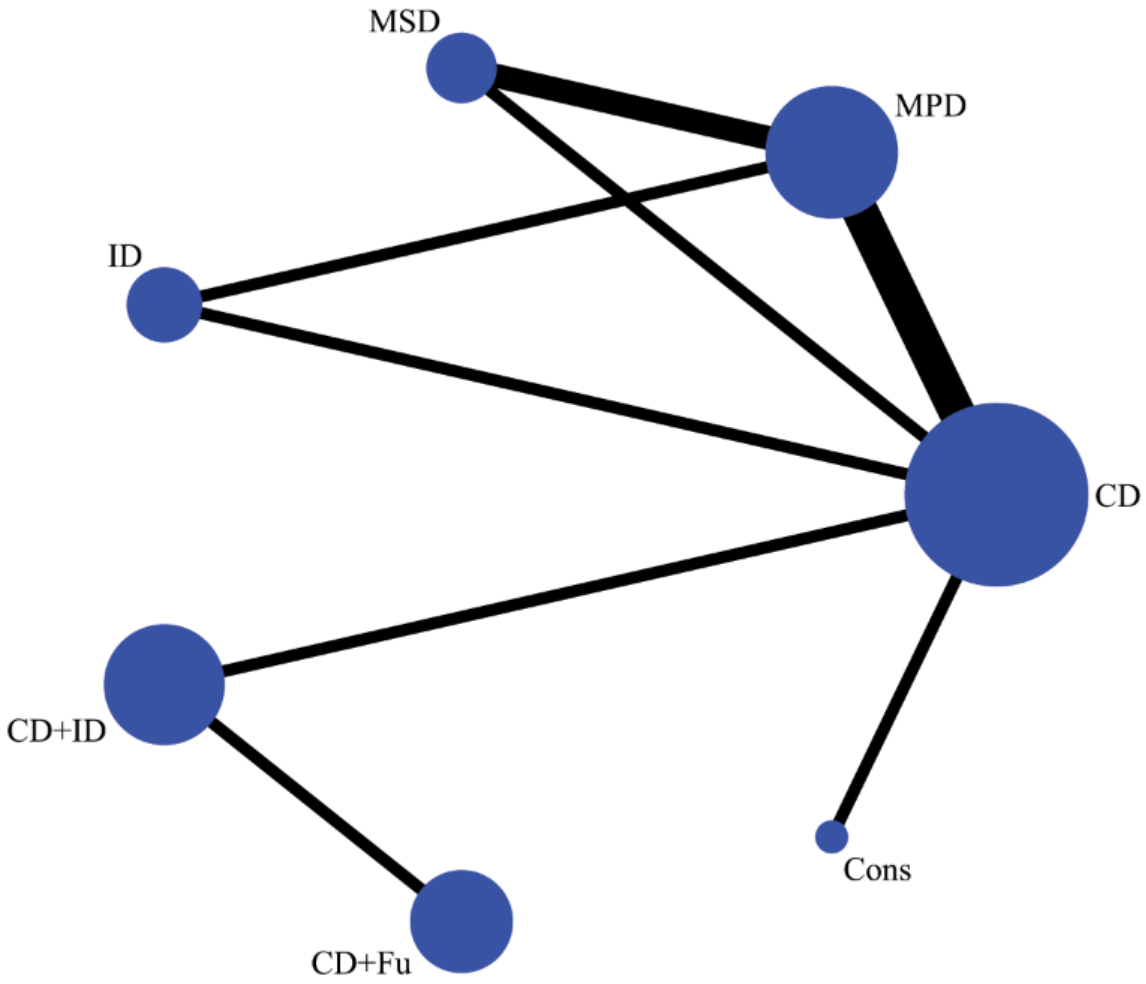
Long-term physical function



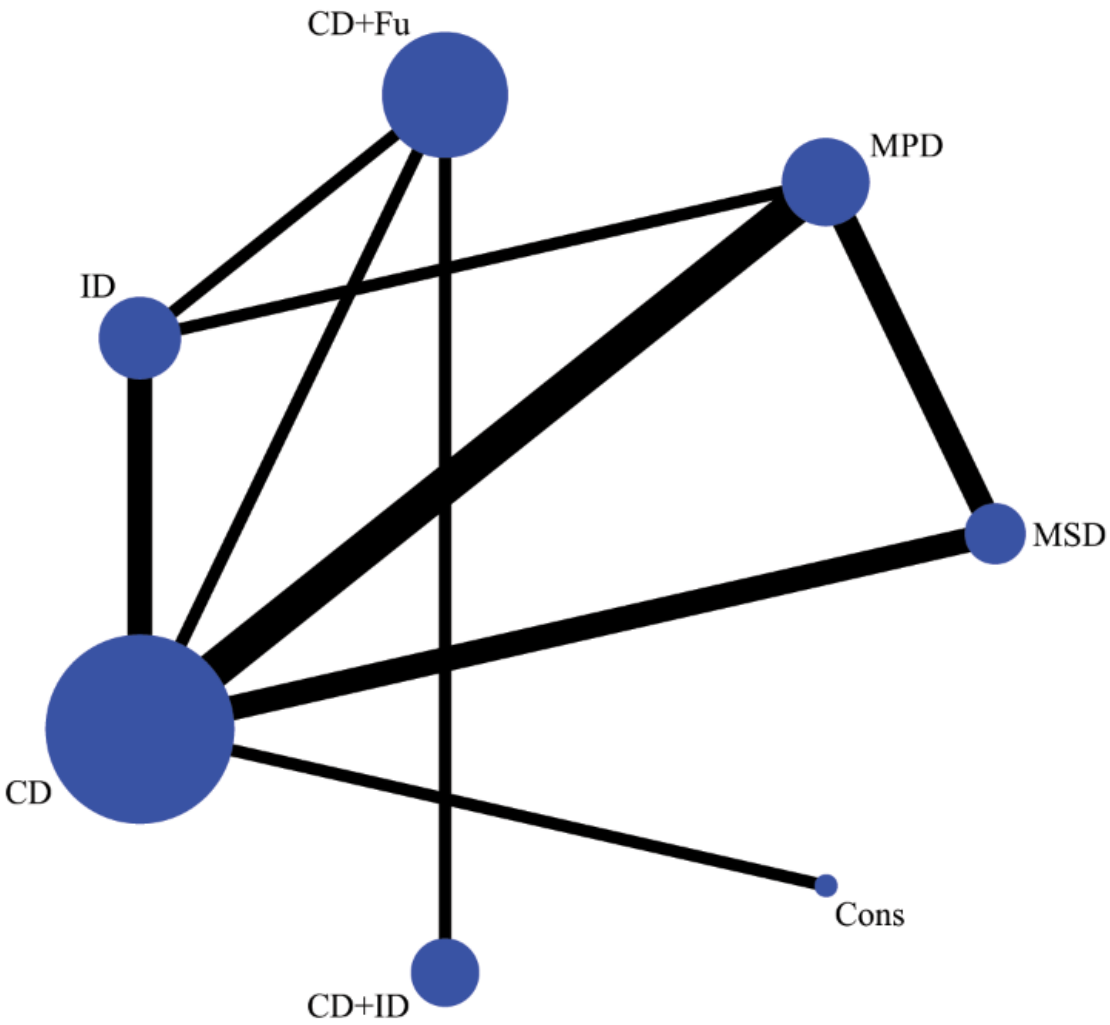
All-cause mortality



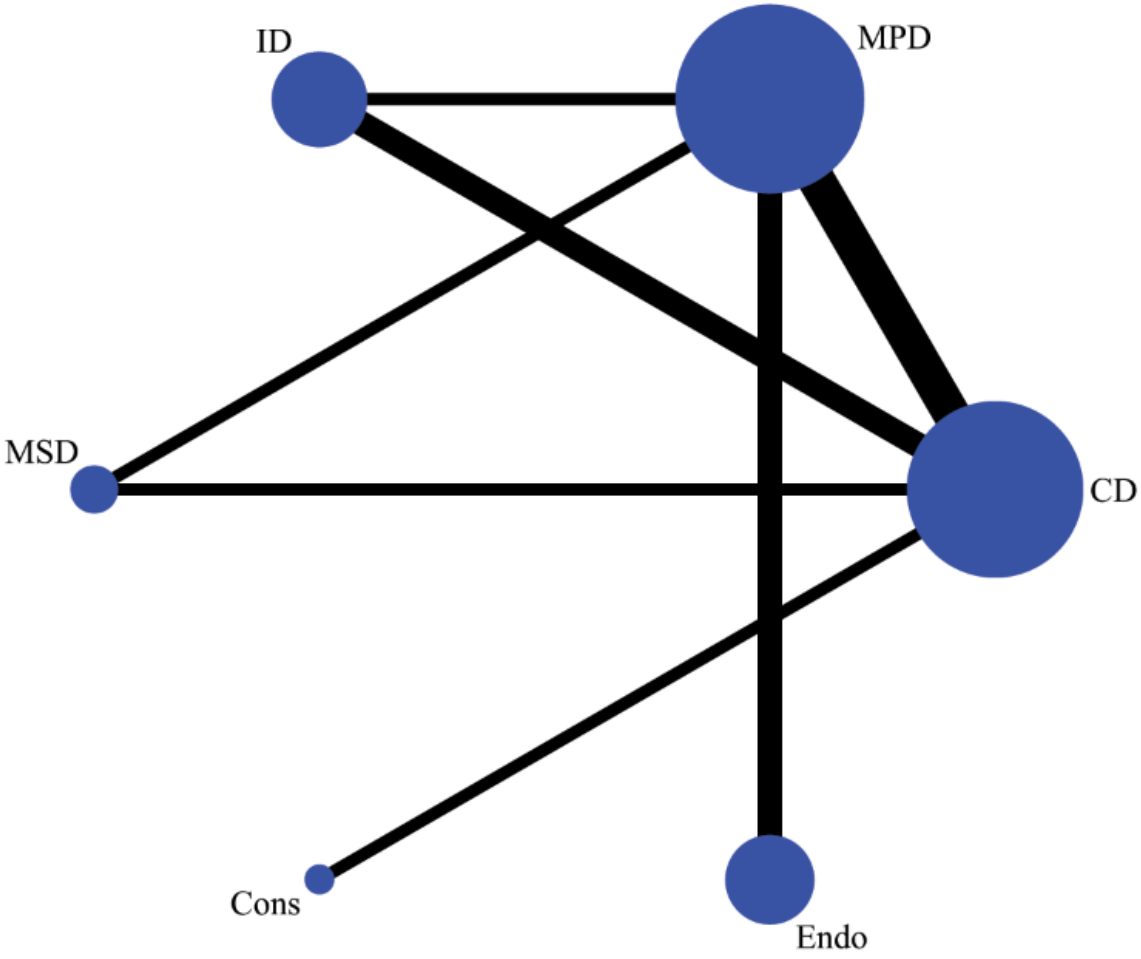
Short-term back pain



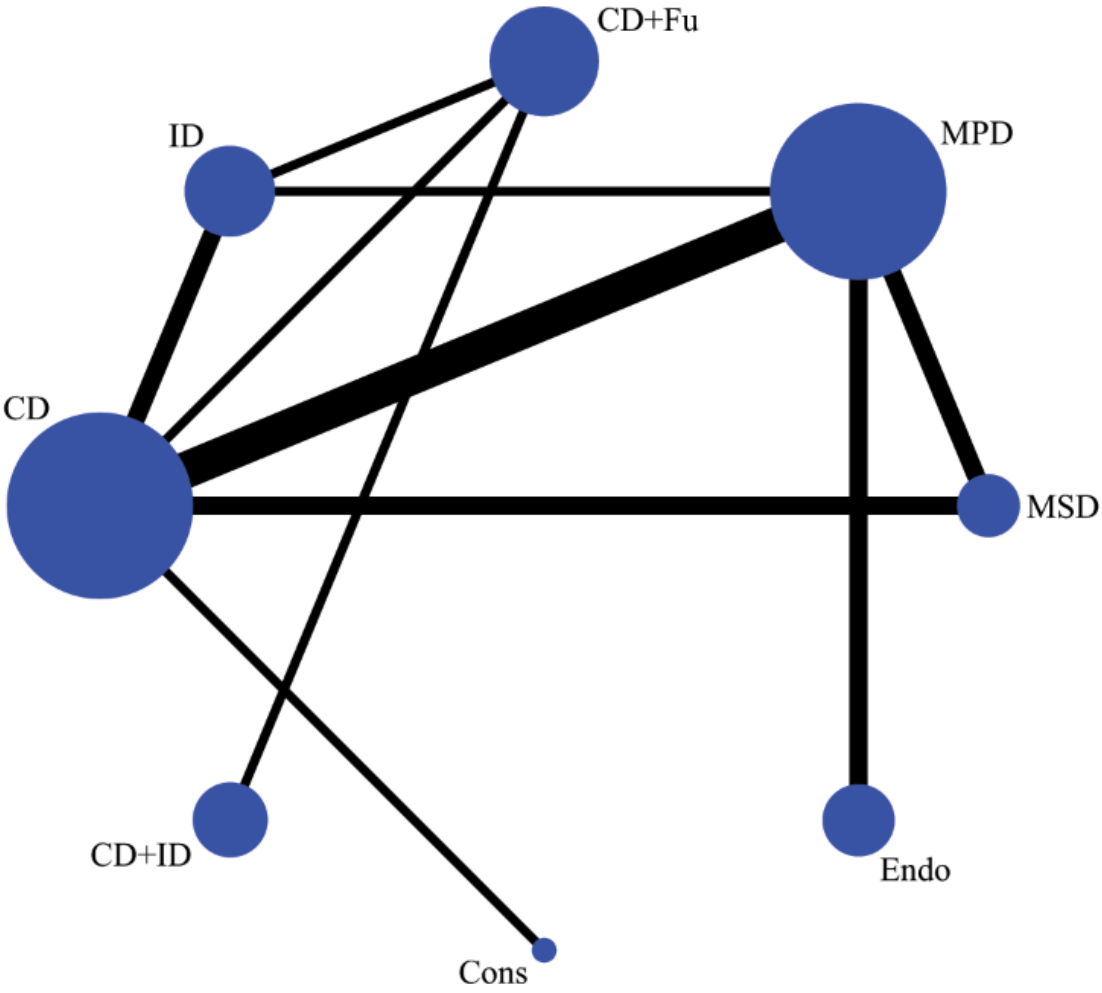
Long-term back pain



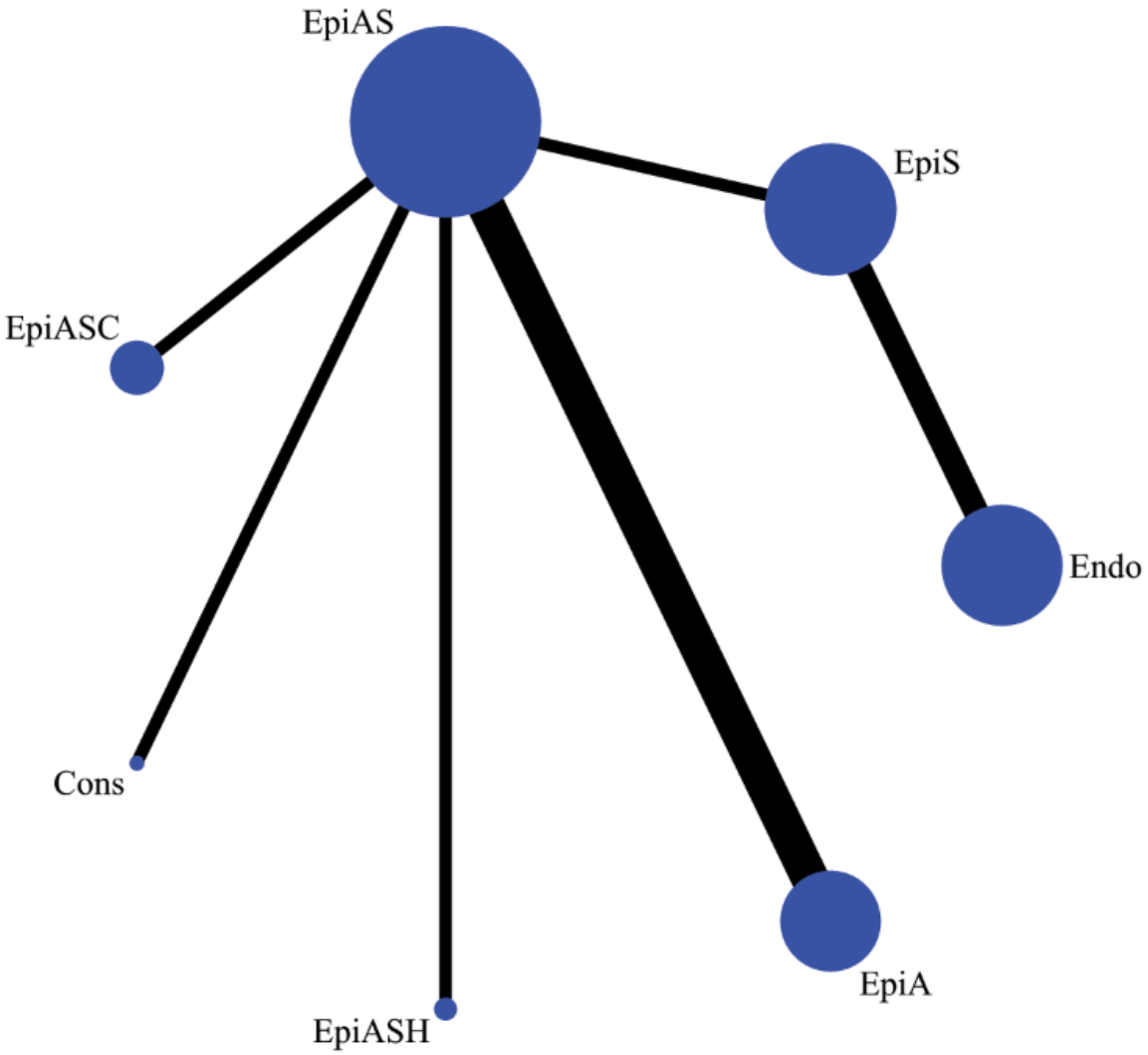
Short-term leg pain



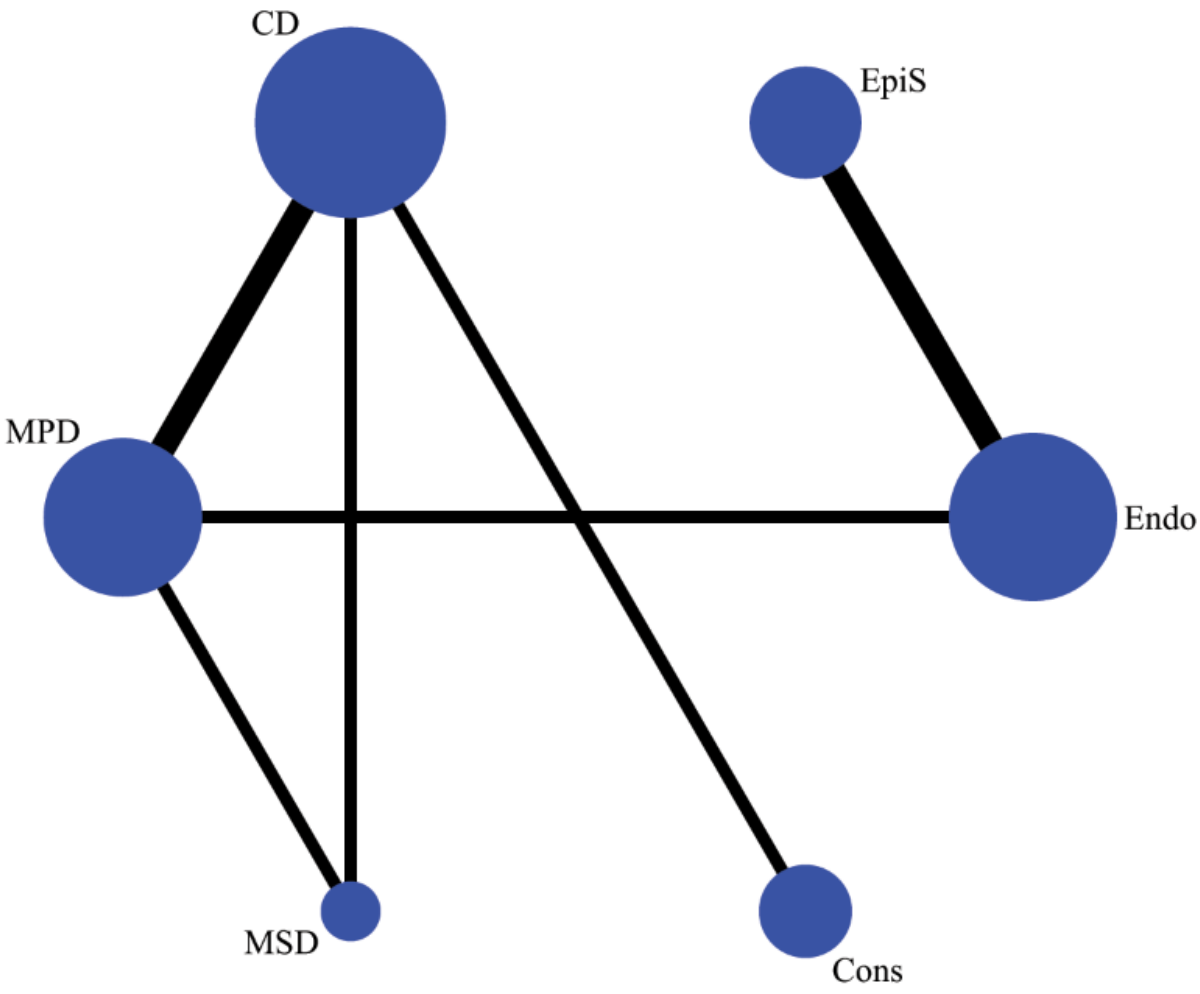
Long-term leg pain



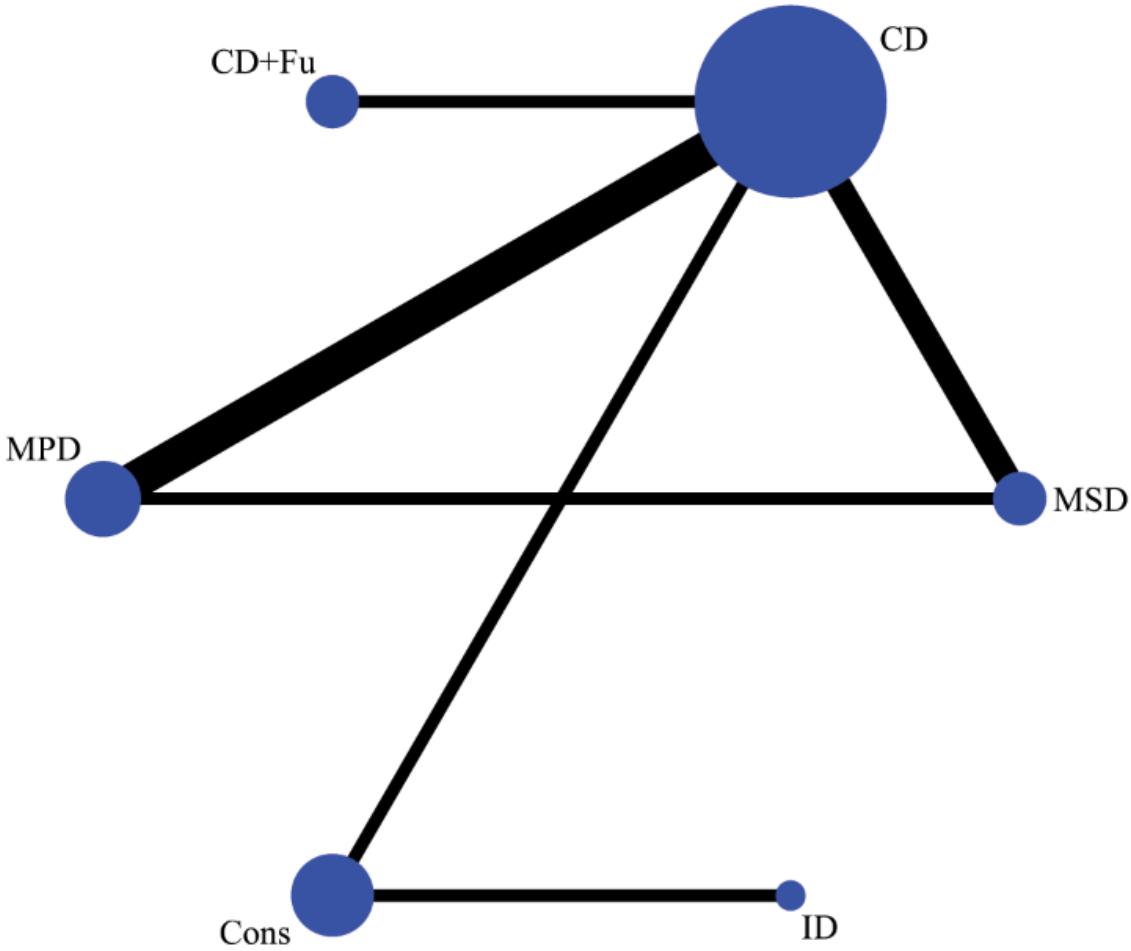
Short-term overall pain



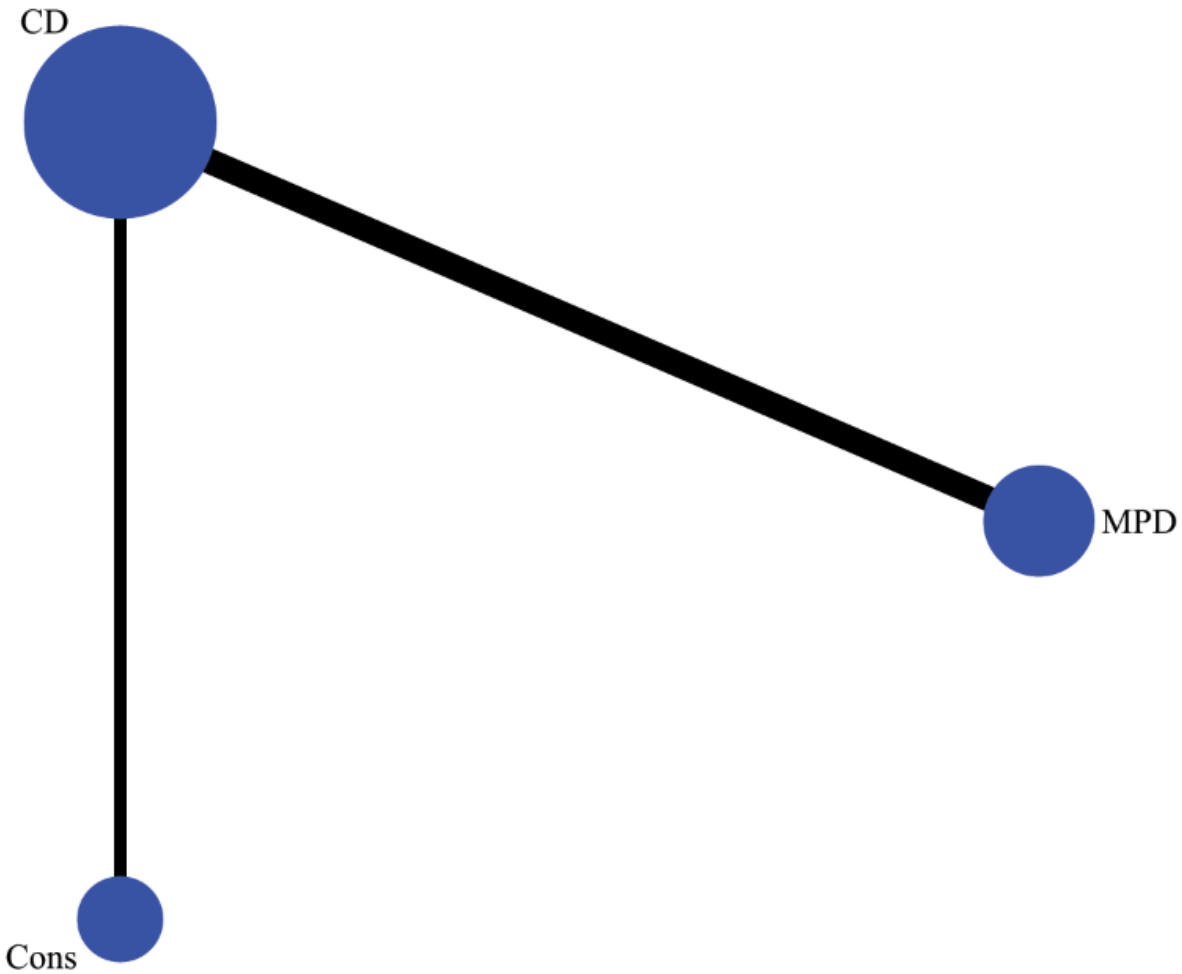
Short-term global impression of recovery



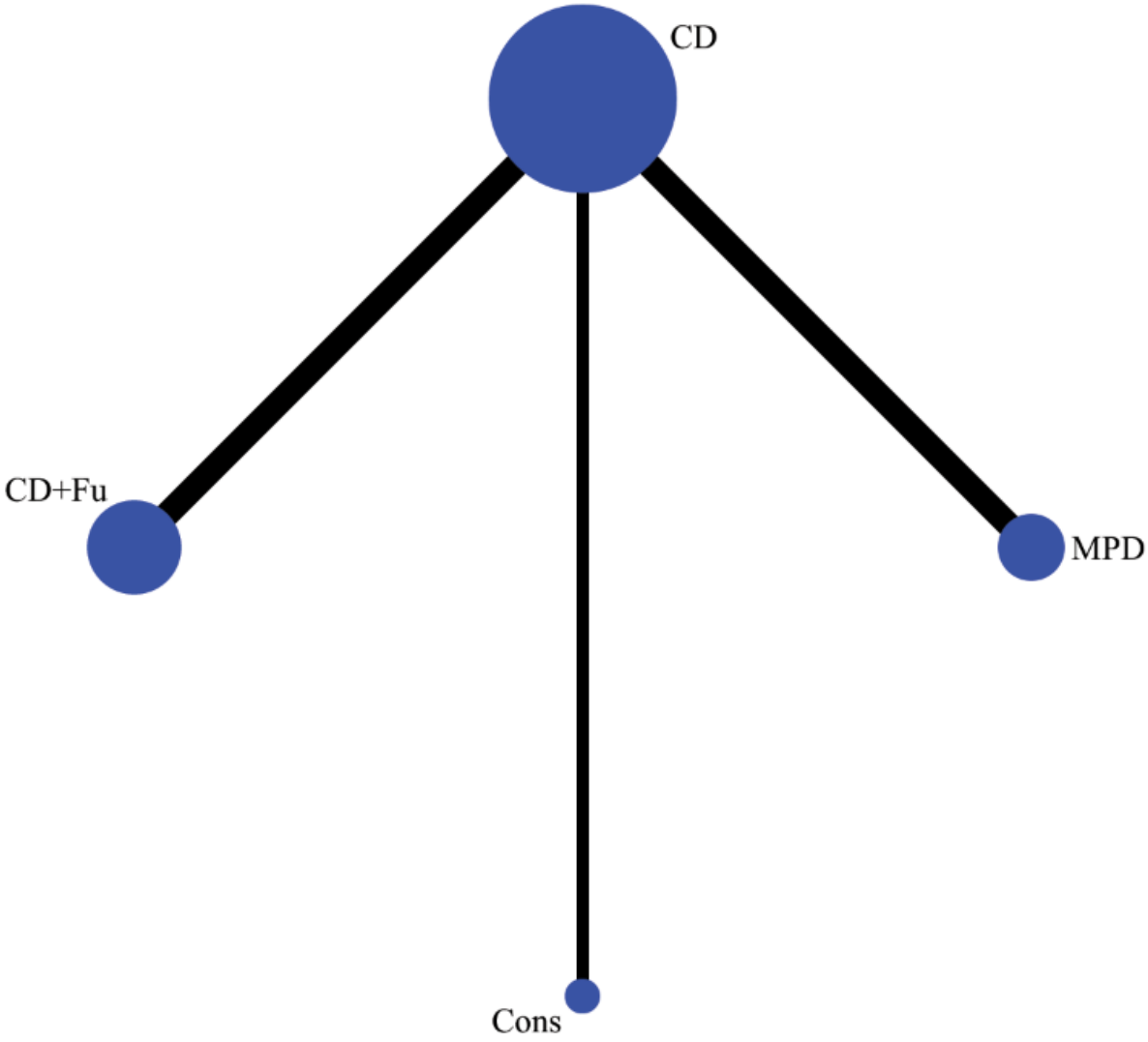
Long-term global impression of recovery



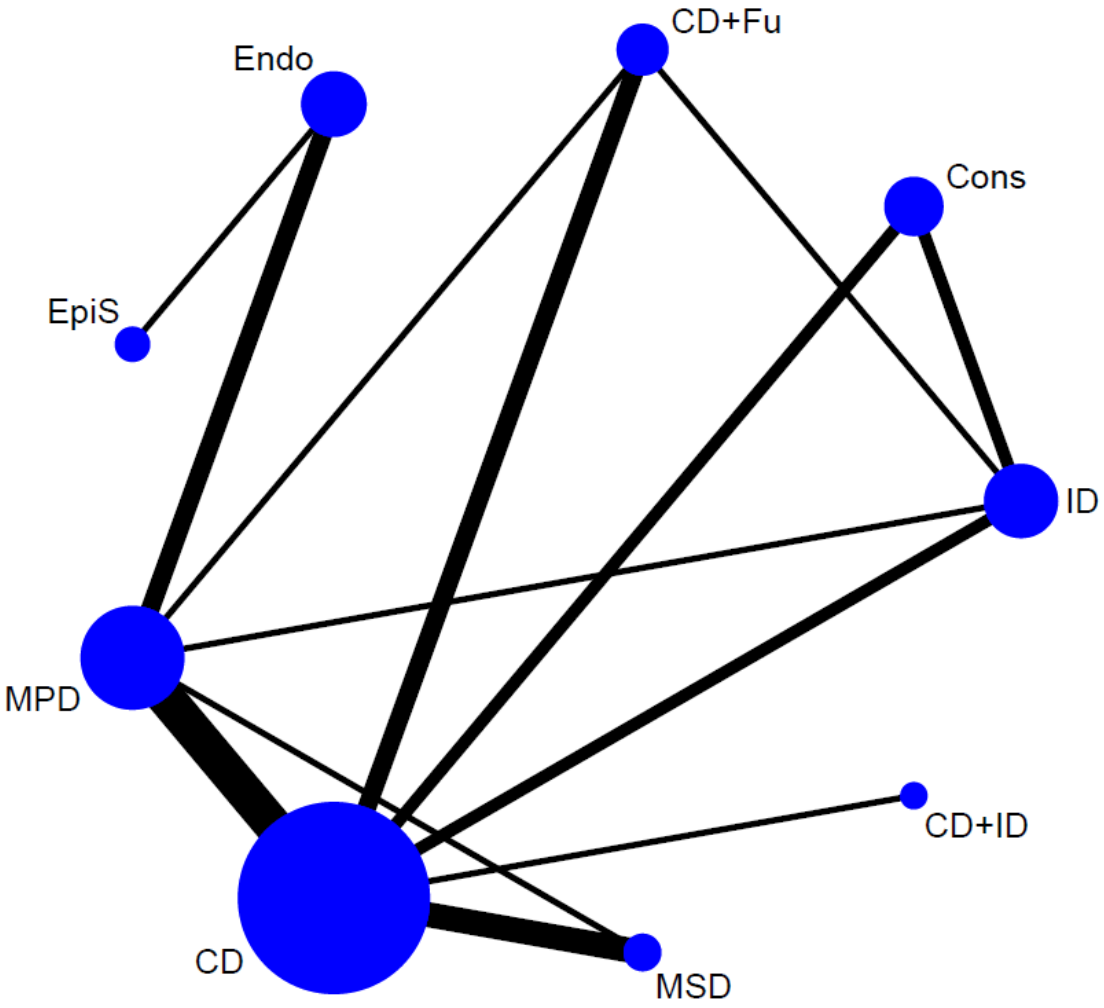
Short-term mobility



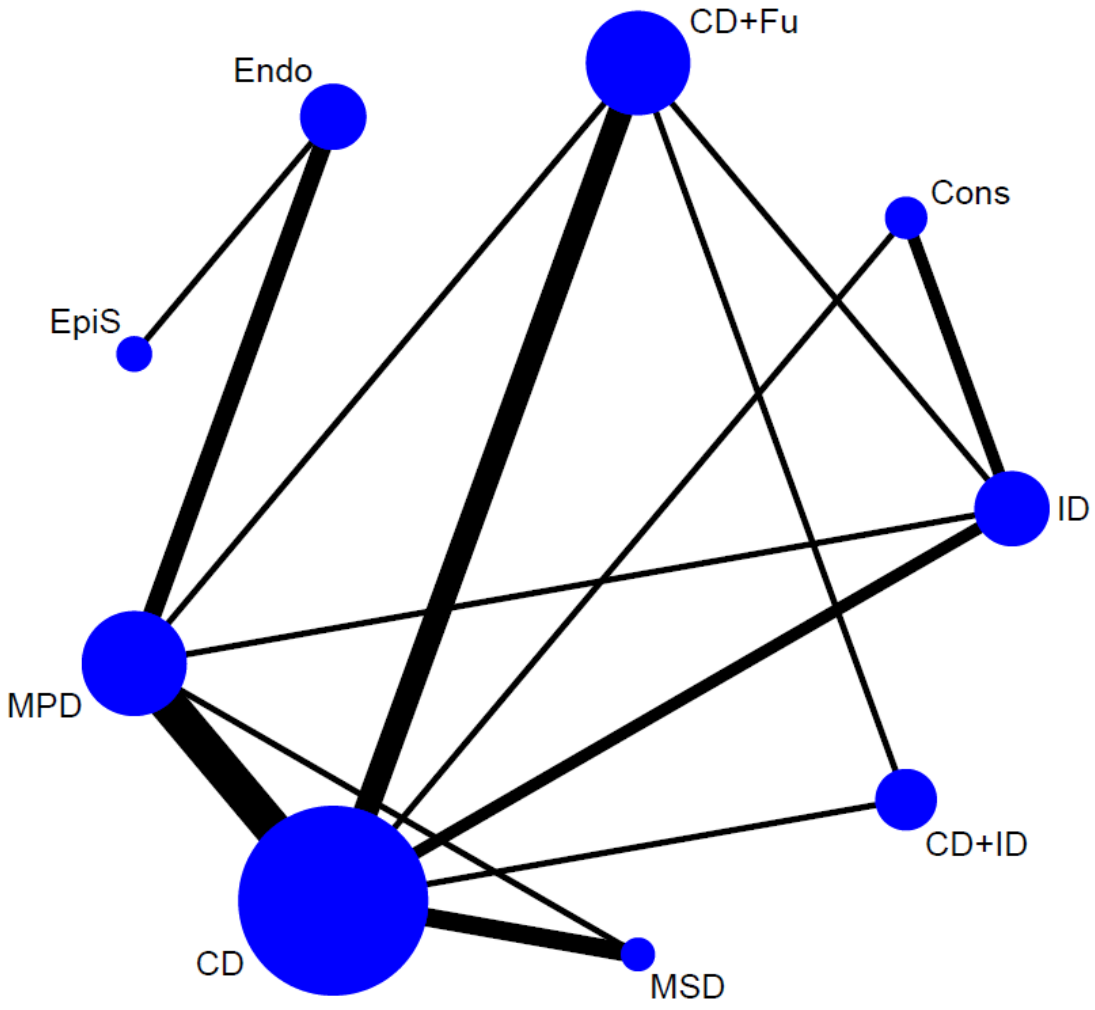
Long-term mobility



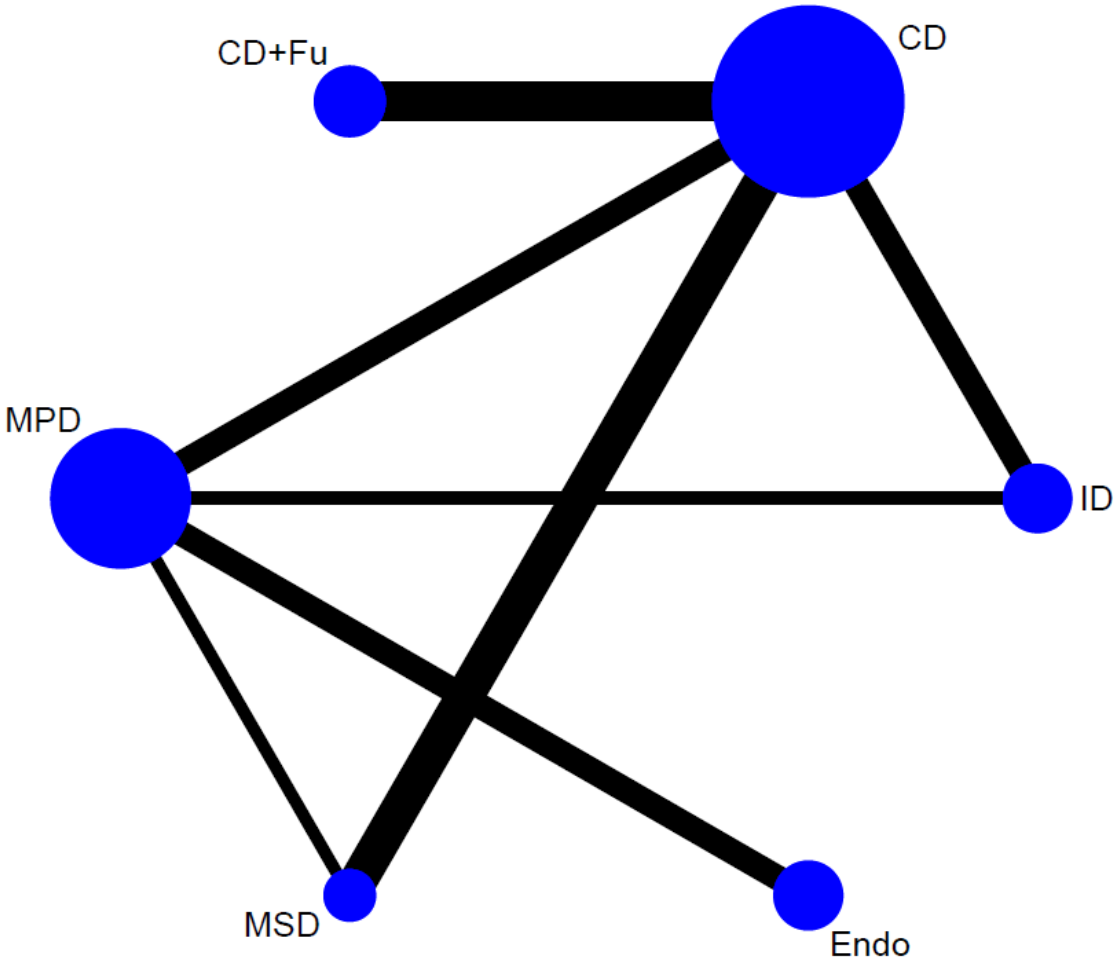
Adverse effect due to any reason



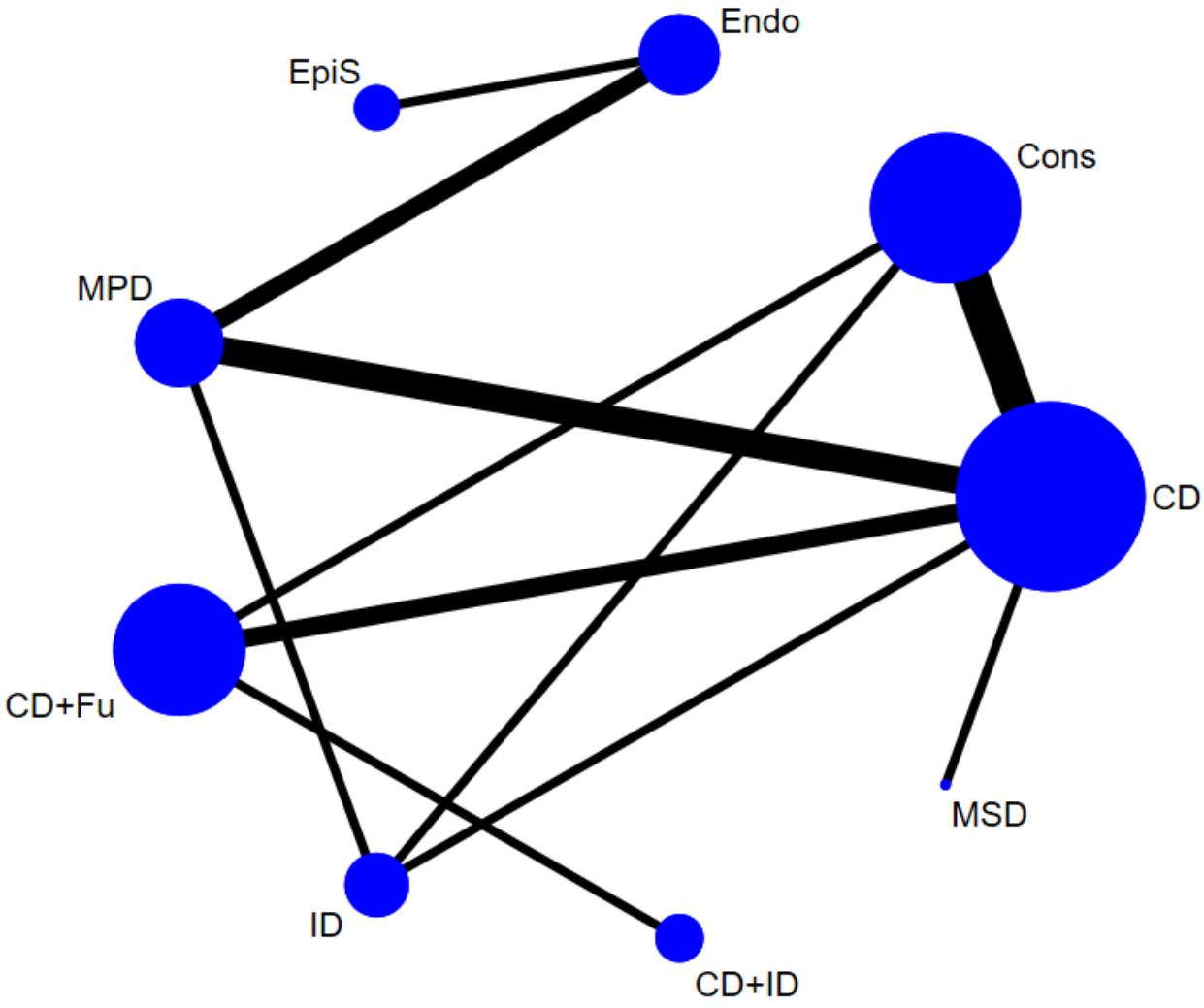
Intervention related adverse effect



Reoperation rate



Treatment withdrawal due to any reason



Appendix S10: GRADE results of primary outcomes

The judgement rule here is: with major concerns in one domain, the confidence rating was degraded one level; with some concerns in one domain, the confidence rating was degraded 0.5 level. If the confidence rating was degraded 1.5 levels, we judged it as 1 level.

Short-term physical function

Comparison	Number of studies	Within-study bias	Reporting bias	Indirectness	Imprecision	Heterogeneity	Incoherence	Confidence rating
BT:EpiAS	1	No concerns	Undetected	No concerns	No concerns	Some concerns	Major concerns	Moderate
BT:EpiASH	1	Some concerns	Undetected	No concerns	Major concerns	No concerns	Major concerns	Low
CD:CD+Fu	1	Major concerns	Undetected	No concerns	Major concerns	No concerns	Major concerns	Very low
CD:Cons	3	Some concerns	Undetected	No concerns	Some concerns	Some concerns	No concerns	Moderate
CD:ID	2	Major concerns	Undetected	No concerns	Some concerns	Some concerns	Major concerns	Very low
CD:MPD	4	Major concerns	Undetected	No concerns	Some concerns	Some concerns	Major concerns	Very low
CD:MSD	1	Major concerns	Undetected	No concerns	Major concerns	No concerns	No concerns	Low
CD+Fu:CD+ID	1	Some concerns	Undetected	No concerns	Major concerns	No concerns	Major concerns	Low
Cons:EpiAS	1	Some concerns	Undetected	No concerns	Some concerns	Some concerns	No concerns	Moderate
Endo:EpiS	2	Major concerns	Undetected	No concerns	No concerns	Major concerns	No concerns	Low
Endo:MPD	2	Some concerns	Undetected	No concerns	Some concerns	Some concerns	No concerns	Moderate
EpiA:EpiAS	4	No concerns	Undetected	No concerns	Some concerns	Some concerns	Major concerns	Low
EpiAS:EpiASH	1	No concerns	Undetected	No concerns	No concerns	Some concerns	Major concerns	Moderate
EpiAS:EpiS	1	Major concerns	Undetected	No concerns	Major concerns	No concerns	No concerns	Low

EpiS:ID	1	Major concerns	Undetected	No concerns	No concerns	No concerns	Some concerns	Moderate
ID:MPD	1	Major concerns	Undetected	No concerns	Some concerns	Some concerns	Some concerns	Low
MPD:MSD	1	Major concerns	Undetected	No concerns	Major concerns	No concerns	No concerns	Low
BT:CD	0	Some concerns	Undetected	No concerns	Major concerns	No concerns	Major concerns	Low
BT:CD+Fu	0	Major concerns	Undetected	No concerns	Major concerns	No concerns	Major concerns	Very low
BT:CD+ID	0	Some concerns	Undetected	No concerns	Major concerns	No concerns	Major concerns	Low
BT:Cons	0	Some concerns	Undetected	No concerns	Major concerns	No concerns	Major concerns	Low
BT:Endo	0	Major concerns	Undetected	No concerns	Major concerns	No concerns	Major concerns	Very low
BT:EpiA	0	No concerns	Undetected	No concerns	Some concerns	Some concerns	Major concerns	Low
BT:EpiS	0	Major concerns	Undetected	No concerns	Some concerns	Some concerns	Major concerns	Very low
BT:ID	0	Major concerns	Undetected	No concerns	Major concerns	No concerns	Major concerns	Very low
BT:MPD	0	Some concerns	Undetected	No concerns	Major concerns	No concerns	Major concerns	Low
BT:MSD	0	Major concerns	Undetected	No concerns	Major concerns	No concerns	Major concerns	Very low
CD:CD+ID	0	Some concerns	Undetected	No concerns	Major concerns	No concerns	Major concerns	Low
CD:Endo	0	Some concerns	Undetected	No concerns	Some concerns	Some concerns	Major concerns	Low
CD:EpiA	0	Major concerns	Undetected	No concerns	Some concerns	Some concerns	Major concerns	Very low
CD:EpiAS	0	Some concerns	Undetected	No concerns	No concerns	Major concerns	Major concerns	Low
CD:EpiASH	0	Some concerns	Undetected	No concerns	Major concerns	No concerns	Major concerns	Low
CD:EpiS	0	Major concerns	Undetected	No concerns	No concerns	Some concerns	Major concerns	Low
CD+Fu:Cons	0	Major concerns	Undetected	No concerns	Major concerns	No concerns	Major concerns	Very low

CD+Fu:Endo	0	Major concerns	Undetected	No concerns	Major concerns	No concerns	Major concerns	Very low
CD+Fu:EpiA	0	Major concerns	Undetected	No concerns	Some concerns	Some concerns	Major concerns	Very low
CD+Fu:EpiAS	0	Major concerns	Undetected	No concerns	Some concerns	Some concerns	Major concerns	Very low
CD+Fu:EpiASH	0	Major concerns	Undetected	No concerns	Major concerns	No concerns	Major concerns	Very low
CD+Fu:EpiS	0	Major concerns	Undetected	No concerns	Some concerns	Some concerns	Major concerns	Very low
CD+Fu:ID	0	Major concerns	Undetected	No concerns	Major concerns	No concerns	Major concerns	Very low
CD+Fu:MPD	0	Major concerns	Undetected	No concerns	Major concerns	No concerns	Major concerns	Very low
CD+Fu:MSD	0	Major concerns	Undetected	No concerns	Major concerns	No concerns	Major concerns	Very low
CD+ID:Cons	0	Some concerns	Undetected	No concerns	Major concerns	No concerns	Major concerns	Low
CD+ID:Endo	0	Some concerns	Undetected	No concerns	Major concerns	No concerns	Major concerns	Low
CD+ID:EpiA	0	Major concerns	Undetected	No concerns	Some concerns	Some concerns	Major concerns	Very low
CD+ID:EpiAS	0	Some concerns	Undetected	No concerns	Some concerns	Some concerns	Major concerns	Low
CD+ID:EpiASH	0	Some concerns	Undetected	No concerns	Major concerns	No concerns	Major concerns	Low
CD+ID:EpiS	0	Major concerns	Undetected	No concerns	Some concerns	Some concerns	Major concerns	Very low
CD+ID:ID	0	Major concerns	Undetected	No concerns	Major concerns	No concerns	Major concerns	Very low
CD+ID:MPD	0	Major concerns	Undetected	No concerns	Major concerns	No concerns	Major concerns	Very low
CD+ID:MSD	0	Major concerns	Undetected	No concerns	Major concerns	No concerns	Major concerns	Very low
Cons:Endo	0	Some concerns	Undetected	No concerns	Major concerns	No concerns	Major concerns	Low
Cons:EpiA	0	Some concerns	Undetected	No concerns	Some concerns	Some concerns	Major concerns	Low
Cons:EpiASH	0	Some concerns	Undetected	No concerns	Major concerns	No concerns	Major concerns	Low

Cons:EpiS	0	Major concerns	Undetected	No concerns	Some concerns	Some concerns	Major concerns	Very low
Cons:ID	0	Major concerns	Undetected	No concerns	Some concerns	Some concerns	Major concerns	Very low
Cons:MPD	0	Major concerns	Undetected	No concerns	Major concerns	No concerns	Major concerns	Very low
Cons:MSD	0	Major concerns	Undetected	No concerns	Major concerns	No concerns	Major concerns	Very low
Endo:EpiA	0	Major concerns	Undetected	No concerns	Major concerns	No concerns	Major concerns	Very low
Endo:EpiAS	0	Major concerns	Undetected	No concerns	Some concerns	Some concerns	Major concerns	Very low
Endo:EpiASH	0	Major concerns	Undetected	No concerns	Major concerns	No concerns	Major concerns	Very low
Endo:ID	0	Major concerns	Undetected	No concerns	Some concerns	Some concerns	Major concerns	Very low
Endo:MSD	0	Major concerns	Undetected	No concerns	Major concerns	No concerns	Major concerns	Very low
EpiA:EpiASH	0	No concerns	Undetected	No concerns	Some concerns	Some concerns	Major concerns	Low
EpiA:EpiS	0	Major concerns	Undetected	No concerns	Major concerns	No concerns	Major concerns	Very low
EpiA:ID	0	Major concerns	Undetected	No concerns	No concerns	Some concerns	Major concerns	Low
EpiA:MPD	0	Major concerns	Undetected	No concerns	Some concerns	Some concerns	Major concerns	Very low
EpiA:MSD	0	Major concerns	Undetected	No concerns	Major concerns	No concerns	Major concerns	Very low
EpiAS:ID	0	Major concerns	Undetected	No concerns	No concerns	Some concerns	Major concerns	Low
EpiAS:MPD	0	Major concerns	Undetected	No concerns	Some concerns	Some concerns	Major concerns	Very low
EpiAS:MSD	0	Major concerns	Undetected	No concerns	Some concerns	Some concerns	Major concerns	Very low
EpiASH:EpiS	0	Major concerns	Undetected	No concerns	Some concerns	Some concerns	Major concerns	Very low
EpiASH:ID	0	Major concerns	Undetected	No concerns	Major concerns	No concerns	Major concerns	Very low
EpiASH:MPD	0	Some concerns	Undetected	No concerns	Major concerns	No concerns	Major concerns	Low

EpiASH:MSD	0	Major concerns	Undetected	No concerns	Major concerns	No concerns	Major concerns	Very low
EpiS:MPD	0	Major concerns	Undetected	No concerns	No concerns	Some concerns	Major concerns	Low
EpiS:MSD	0	Major concerns	Undetected	No concerns	Some concerns	Some concerns	Major concerns	Very low
ID:MSD	0	Major concerns	Undetected	No concerns	Major concerns	No concerns	Major concerns	Very low

Long-term physical function

Comparison	Number of studies	Within-study bias	Reporting bias	Indirectness	Imprecision	Heterogeneity	Incoherence	Confidence rating
CD:CD+Fu	2	Major concerns	Undetected	No concerns	Some concerns	Some concerns	No concerns	Low
CD:CD+ID	2	Major concerns	Undetected	No concerns	Some concerns	Some concerns	No concerns	Low
CD:Cons	3	Some concerns	Undetected	No concerns	Some concerns	Some concerns	Major concerns	Low
CD:ID	2	Major concerns	Undetected	No concerns	Some concerns	Some concerns	Major concerns	Very low
CD:MPD	5	Major concerns	Undetected	No concerns	Some concerns	Some concerns	Major concerns	Very low
CD:MSD	4	Major concerns	Undetected	No concerns	Some concerns	Some concerns	No concerns	Low
CD+Fu:CD+ID	1	Some concerns	Undetected	No concerns	Major concerns	No concerns	No concerns	Moderate
CD+Fu:ID	1	Some concerns	Undetected	No concerns	Some concerns	Some concerns	No concerns	Moderate
CD+Fu:MPD	1	Major concerns	Undetected	No concerns	Some concerns	Some concerns	No concerns	Low
Endo:EpiS	1	Major concerns	Undetected	No concerns	No concerns	Some concerns	No concerns	Moderate
Endo:MPD	2	Some concerns	Undetected	No concerns	Some concerns	Some concerns	No concerns	Moderate
EpiS:ID	2	Major concerns	Undetected	No concerns	No concerns	No concerns	No concerns	Moderate
ID:MPD	1	Major concerns	Undetected	No concerns	Some concerns	Some concerns	Major concerns	Very low
MPD:MSD	2	Major concerns	Undetected	No concerns	Some concerns	Some concerns	No concerns	Low
CD:Endo	0	Major concerns	Undetected	No concerns	Some concerns	Some concerns	Major concerns	Very low
CD:EpiS	0	Major concerns	Undetected	No concerns	No concerns	No concerns	Major concerns	Low
CD+Fu:Cons	0	Major concerns	Undetected	No concerns	Some concerns	Some concerns	Major concerns	Very low
CD+Fu:Endo	0	Major concerns	Undetected	No concerns	Major concerns	No concerns	Major concerns	Very low

CD+Fu:EpiS	0	Major concerns	Undetected	No concerns	No concerns	Some concerns	Major concerns	Low
CD+Fu:MSD	0	Major concerns	Undetected	No concerns	Some concerns	Some concerns	Major concerns	Very low
CD+ID:Cons	0	Major concerns	Undetected	No concerns	Major concerns	No concerns	Major concerns	Very low
CD+ID:Endo	0	Major concerns	Undetected	No concerns	Major concerns	No concerns	Major concerns	Very low
CD+ID:EpiS	0	Major concerns	Undetected	No concerns	No concerns	Some concerns	Major concerns	Low
CD+ID:ID	0	Major concerns	Undetected	No concerns	Some concerns	Some concerns	Major concerns	Very low
CD+ID:MPD	0	Major concerns	Undetected	No concerns	Major concerns	No concerns	Major concerns	Very low
CD+ID:MSD	0	Major concerns	Undetected	No concerns	Some concerns	Some concerns	Major concerns	Very low
Cons:Endo	0	Major concerns	Undetected	No concerns	Major concerns	No concerns	Major concerns	Very low
Cons:EpiS	0	Major concerns	Undetected	No concerns	No concerns	Some concerns	Major concerns	Low
Cons:ID	0	Major concerns	Undetected	No concerns	Some concerns	Some concerns	Major concerns	Very low
Cons:MPD	0	Major concerns	Undetected	No concerns	Major concerns	No concerns	Major concerns	Very low
Cons:MSD	0	Major concerns	Undetected	No concerns	Some concerns	Some concerns	Major concerns	Very low
Endo:ID	0	Major concerns	Undetected	No concerns	Some concerns	Some concerns	Major concerns	Very low
Endo:MSD	0	Major concerns	Undetected	No concerns	Some concerns	Some concerns	Major concerns	Very low
EpiS:MPD	0	Major concerns	Undetected	No concerns	No concerns	Some concerns	Major concerns	Low
EpiS:MSD	0	Major concerns	Undetected	No concerns	No concerns	No concerns	Major concerns	Low
ID:MSD	0	Major concerns	Undetected	No concerns	Major concerns	No concerns	Major concerns	Very low

All-cause mortality

Comparison	Number of studies	Within-study bias	Reporting bias	Indirectness	Imprecision	Heterogeneity	Incoherence	Confidence rating
CD:CD+Fu	2	Some concerns	Undetected	No concerns	No concerns	Major concerns	No concerns	Moderate
CD:CD+ID	1	Major concerns	Undetected	No concerns	Major concerns	No concerns	No concerns	Low
CD:Cons	4	Some concerns	Undetected	No concerns	Major concerns	No concerns	No concerns	Moderate
CD:MPD	1	Major concerns	Undetected	No concerns	Major concerns	No concerns	No concerns	Low
CD:MSD	1	Major concerns	Undetected	No concerns	Major concerns	No concerns	No concerns	Low
CD+Fu:Cons	1	Some concerns	Undetected	No concerns	Major concerns	No concerns	No concerns	Moderate
CD+ID:MPD	1	Some concerns	Undetected	No concerns	Major concerns	No concerns	No concerns	Moderate
MPD:MSD	1	Major concerns	Undetected	No concerns	Major concerns	No concerns	No concerns	Low
CD+Fu:CD+ID	0	Major concerns	Undetected	No concerns	Major concerns	No concerns	No concerns	Low
CD+Fu:MPD	0	Major concerns	Undetected	No concerns	Major concerns	No concerns	No concerns	Low
CD+Fu:MSD	0	Major concerns	Undetected	No concerns	Major concerns	No concerns	No concerns	Low
CD+ID:Cons	0	Major concerns	Undetected	No concerns	Major concerns	No concerns	No concerns	Low
CD+ID:MSD	0	Major concerns	Undetected	No concerns	Major concerns	No concerns	No concerns	Low
Cons:MPD	0	Major concerns	Undetected	No concerns	Major concerns	No concerns	No concerns	Low
Cons:MSD	0	Major concerns	Undetected	No concerns	Major concerns	No concerns	No concerns	Low

Appendix S11: Results from pairwise and network meta-analyses

Long-term physical function (the numbers are presented as standardised mean difference and 95% confidence interval)

Cons			0.29 (-0.38,0.96)					
0.54 (-0.34,1.41)	MSD	-0.1 (-0.98,0.78)	-0.22 (-0.86,0.42)					
0.28 (-0.53,1.09)	-0.26 (-0.88,0.36)	MPD	-0.34 (-0.83,0.15)	-0.03 (-0.86,0.8)		0.08 (-1.21,1.38)		**1.79 (0.75, 2.8)**
0.29 (-0.38,0.96)	-0.24 (-0.80,0.31)	0.01 (-0.44,0.47)	CD		-0.26 (-1.12,0.6)	-0.02 (-0.88,0.85)	-0.72 (-1.87,0.44)	-0.27 (-1,0.45)
0.10 (-0.96,1.16)	-0.44 (-1.37,0.50)	-0.18 (-0.91,0.55)	-0.19 (-1.01,0.63)	Endo				
0.18 (-0.80,1.15)	-0.36 (-1.25,0.53)	-0.10 (-0.92,0.72)	-0.12 (-0.82,0.59)	0.08 (-0.99,1.14)	CD+ID	-0.15 (-1.32,1.01)		
0.28 (-0.61,1.17)	-0.25 (-1.04,0.53)	0.00 (-0.66,0.66)	-0.01 (-0.60,0.57)	0.18 (-0.76,1.12)	0.10 (-0.67,0.88)	CD+Fu		0.17 (-1.08,1.42)
-0.90 (-2.00,0.21)	**-1.43 (-2.44,-0.43)**	**-1.18 (-2.04,-0.31)**	**-1.19 (-2.07,-0.31)**	**-1.00 (-1.87,-0.12)**	-1.07 (-2.18,0.03)	**-1.18 (-2.15,-0.21)**	EpiS	**1.74 (0.88, 2.6)**
0.69 (-0.21,1.59)	0.15 (-0.64,0.94)	0.41 (-0.23,1.05)	0.40 (-0.21,1.00)	0.59 (-0.27,1.45)	0.51 (-0.38,1.41)	0.41 (-0.32,1.13)	**1.59 (0.84,2.33)**	ID

The results from bottom-left are network evidence. The results from top-right are pairwise evidence. For the network evidence, the minus sign means the up to the left intervention has the superior result. For the pairwise evidence, the minus sign means the leftward intervention has the superior result. The statistically significant results were presented in bold.

All-cause mortality (the numbers are presented as odds ratio and 95% confidence interval)

Cons				2.64 (0.67,10.40)	0.82 (0.34, 1.97)
0.24 (0.01,6.72)	CD+ID				3.16 (0.13,78.51)
0.25 (0.01,4.40)	1.02 (0.01,70.13)	MSD	1.00 (0.11,9.30)		3.09 (0.20,48.23)
0.25 (0.02,2.84)	1.02 (0.02,52.78)	1.00 (0.11,9.30)	MPD		3.09 (0.31,30.54)
3.05 (0.94,9.92)	12.51 (0.39,399.39)	12.22 (0.59,255.00)	12.24 (0.88,169.94)	CD+Fu	0.18 (0.02,1.51)

0.77 (0.34,1.77)	3.16 (0.13,78.51)	3.09 (0.20,48.23)	3.09 (0.31,30.54)	**0.25** **(0.07,0.92)**	CD

The results from bottom-left are network evidence. The results from top-right are pairwise evidence. For the network evidence, the value less than one means the up to the left intervention has the superior result. For the pairwise evidence, the minus sign means the leftward intervention has the superior result.

Short-term back pain (the numbers are presented as mean difference and 95% confidence interval)

Cons						2.5 (-0.91,5.91)
1.10 (-4.79,6.99)	CD+Fu	0.60 (-2.65,3.85)				
1.70 (-3.21,6.61)	0.60 (-2.65,3.85)	CD+ID				0.8 (-2.73,4.33)
2.89 (-1.31,7.09)	1.79 (-3.61,7.19)	1.19 (-3.11,5.50)	ID		-0.34 (-3.93,3.25)	0.1 (-3.57,3.77)
2.87 (-1.31,7.06)	1.77 (-3.62,7.16)	1.17 (-3.12,5.47)	-0.02 (-3.15,3.11)	MSD	0.45 (-1.92, 2.82)	-0.1 (-3.77,3.56)
3.02 (-0.79,6.83)	1.92 (-3.19,7.02)	1.32 (-2.61,5.25)	0.13 (-2.32,2.57)	0.15 (-2.01,2.31)	MPD	-1 (-2.78,0.78)
2.50 (-0.90,5.90)	1.40 (-3.41,6.21)	0.80 (-2.74,4.34)	-0.39 (-2.85,2.07)	-0.37 (-2.81,2.06)	-0.52 (-2.24,1.20)	CD

The results from bottom-left are network evidence. The results from top-right are pairwise evidence. For the network evidence, the minus sign means the up to the left intervention has the superior result. For the pairwise evidence, the minus sign means the leftward intervention has the superior result.

Long-term back pain (the numbers are presented as mean difference and 95% confidence interval)

Cons		2.30 (-0.42,5.01)				
1.98 (-2.22,6.18)	CD+ID			-0.10 (-2.60,2.40)		

2.30 (-0.42,5.01)	0.32 (-2.89,3.53)	CD	-0.11 (-2.05,1.83)	-0.4 (-3.12,2.32)	0.6 (-0.87,2.07)	0.44 (-1.48,2.36)
2.65 (-0.42,5.73)	0.68 (-2.57,3.92)	0.36 (-1.09,1.80)	ID	-0.8 (-3.8,2.2)	-0.85 (-3.38,1.68)	
1.88 (-1.50,5.25)	-0.10 (-2.60,2.40)	-0.42 (-2.43,1.59)	-0.78 (-2.83,1.28)	CD+Fu		
2.57 (-0.41,5.55)	0.59 (-2.78,3.96)	0.27 (-0.95,1.49)	-0.08 (-1.70,1.53)	0.69 (-1.57,2.95)	MPD	-0.22 (-1.94,1.5)
2.83 (-0.27,5.93)	0.85 (-2.66,4.36)	0.53 (-0.97,2.04)	0.18 (-1.79,2.14)	0.95 (-1.51,3.41)	0.26 (-1.23,1.75)	MSD

The results from bottom-left are network evidence. The results from top-right are pairwise evidence. For the network evidence, the minus sign means the up to the left intervention has the superior result. For the pairwise evidence, the minus sign means the leftward intervention has the superior result.

Short-term leg pain (the numbers are presented as mean difference and 95% confidence interval)

Cons					**1.80 (0.11,3.49)**
2.03 (-1.41,5.46)	Endo			-0.30 (-3.22,2.62)	
2.47 (0.50,4.45)	0.45 (-2.64,3.53)	MSD		-0.55 (-1.75,0.65)	-0.87 (-2.05,0.31)
2.05 (0.12,3.98)	0.02 (-3.02,3.07)	-0.42 (-1.70,0.85)	ID	**-0.96 (-1.16,-0.76)**	0.43 (-0.28,1.13)
1.73 (-0.09,3.54)	-0.30 (-3.21,2.61)	-0.75 (-1.76,0.27)	-0.32 (-1.22,0.57)	MPD	**-0.29 (-0.49,-0.1)**
1.80 (0.11,3.49)	-0.23 (-3.21,2.76)	-0.67 (-1.69,0.34)	-0.25 (-1.18,0.68)	0.07 (-0.59,0.74)	CD

The results from bottom-left are network evidence. The results from top-right are pairwise evidence. For the network evidence, the minus sign means the up to the left intervention has the superior result. For the pairwise evidence, the minus sign means the leftward intervention has the superior result. The statistically significant results were presented in bold.

Long-term leg pain (the numbers are presented as mean difference and 95% confidence interval)

Cons			1.70 (-0.12,3.52)				
1.65 (-1.90,5.19)	Endo					0.05 (-2.90,3.00)	
1.51 (-1.37,4.39)	-0.14 (-3.89,3.61)	CD+ID			-0.03 (-1.48,1.42)		
1.70 (-0.13,3.53)	0.05 (-2.99,3.09)	0.19 (-2.03,2.42)	CD	-0.03 (-1.48,1.42)	-0.1 (-1.98,1.78)	0.05 (-0.81,0.91)	0.75 (-0.52,2.02)
1.90 (-0.19,3.99)	0.25 (-2.86,3.37)	0.40 (-1.97,2.76)	0.20 (-0.81,1.22)	ID	-1 (-5.02,3.02)	-0.34 (-1.81,1.13)	
1.48 (-1.01,3.97)	-0.17 (-3.63,3.29)	-0.03 (-1.48,1.42)	-0.22 (-1.91,1.47)	-0.43 (-2.30,1.45)	CD+Fu		
1.70 (-0.27,3.67)	0.05 (-2.90,3.00)	0.19 (-2.12,2.51)	0.00 (-0.73,0.73)	-0.20 (-1.21,0.80)	0.22 (-1.58,2.03)	MPD	0.02 (-0.55,0.59)
1.99 (-0.09,4.07)	0.34 (-2.74,3.42)	0.48 (-1.94,2.90)	0.29 (-0.70,1.29)	0.09 (-1.19,1.37)	0.51 (-1.43,2.45)	0.29 (-0.59,1.18)	MSD

The results from bottom-left are network evidence. The results from top-right are pairwise evidence. For the network evidence, the minus sign means the up to the left intervention has the superior result. For the pairwise evidence, the minus sign means the leftward intervention has the superior result.

Short-term overall pain (the numbers are presented as mean difference and 95% confidence interval)

Cons				-0.10 (-2.00,1.80)		
-0.22 (-2.21,1.76)	EpiA			0.12 (-0.46,0.71)		
2.10 (-0.02,4.22)	**2.32 (1.21,3.44)**	EpiASH		**-2.20 (-3.14,-1.26)**		
3.10 (0.64,5.56)	**3.32 (1.66,4.99)**	1.00 (-0.82,2.82)	EpiASC	**-3.20 (-4.76,-1.64)**		
-0.10 (-2.00,1.80)	0.12 (-0.46,0.71)	**-2.20 (-3.14,-1.26)**	**-3.20 (-4.76,-1.64)**	EpiAS	-0.80 (-2.25,0.65)	
-0.90 (-3.29,1.49)	-0.68 (-2.24,0.89)	**-3.00 (-4.73,-1.27)**	**-4.00 (-6.13,-1.87)**	-0.80 (-2.25,0.65)	EpiS	**2.10 (1.20,3.00)**

1.20 (-1.35,3.75)	1.42 (-0.39,3.24)	-0.90 (-2.85,1.05)	-1.90 (-4.21,0.41)	1.30 (-0.41,3.00)	**2.10** **(1.20,3.00)**	Endo

The results from bottom-left are network evidence. The results from top-right are pairwise evidence. For the network evidence, the minus sign means the up to the left intervention has the superior result. For the pairwise evidence, the minus sign means the leftward intervention has the superior result. The statistically significant results were presented in bold.

Short-term global impression of recovery (the numbers are presented as odds ratio and 95% confidence interval)

Cons			0.49 (0.22,1.08)		
0.50 (0.13,1.95)	MSD	0.28 (0.06,1.23)	1.15 (0.36, 3.63)		
0.21 **(0.06,0.71)**	0.42 (0.12,1.49)	MPD	2.27 (0.93,5.57)		0.50 (0.16,1.60)
0.49 (0.22,1.08)	0.96 (0.32,2.87)	2.27 (0.93,5.57)	CD		
0.39 (0.06,2.62)	0.76 (0.11,5.53)	1.81 (0.44,7.45)	0.79 (0.14,4.54)	EpiS	**0.28** **(0.12,0.63)**
0.11 **(0.02,0.57)**	0.21 (0.04,1.18)	0.50 (0.16,1.60)	**0.22** **(0.05,0.95)**	**0.28** **(0.12,0.63)**	Endo

The results from bottom-left are network evidence. The results from top-right are pairwise evidence. For the network evidence, the value larger than one means the up to the left intervention has the superior result. For the pairwise evidence, the minus sign means the leftward intervention has the superior result. The statistically significant results were presented in bold.

Long-term global impression of recovery (the numbers are presented as odds ratio and 95% confidence interval)

Cons	0.42 (0.07,2.45)			0.50 (0.10,2.56)	
0.42 (0.07,2.45)	ID				
0.11 **(0.01,0.89)**	0.27 (0.02,4.00)	MPD		**3.32** **(1.61,6.86)**	**9.78** **(1.58,60.51)**

0.55 (0.06,5.41)	1.29 (0.07,23.09)	4.83 (0.62,37.41)	CD+Fu	0.92 (0.18,4.59)	
0.50 (0.10,2.56)	1.18 (0.11,12.96)	**4.43** **(1.25,15.67)**	0.92 (0.18,4.59)	CD	0.32 (0.1, 1.06)
0.19 (0.02,1.60)	0.46 (0.03,7.12)	1.72 (0.31,9.69)	0.36 (0.04,2.88)	0.39 (0.10,1.48)	MSD

The results from bottom-left are network evidence. The results from top-right are pairwise evidence. For the network evidence, the value larger than one means the up to the left intervention has the superior result. For the pairwise evidence, the minus sign means the leftward intervention has the superior result. The statistically significant results were presented in bold.

Short-term mobility (the numbers are presented as standardised mean difference and 95% confidence interval)

Cons	0 (-0.41, 0.41)	
0 (-0.41, 0.41)	CD	0.23 (-0.12, 0.59)
0.23 (-0.31,0.77)	0.23 (-0.12, 0.59)	MPD

The results from bottom-left are network evidence. The results from top-right are pairwise evidence. For the network evidence, the minus sign means the up to the left intervention has the superior result. For the pairwise evidence, the minus sign means the leftward intervention has the superior result.

Long-term mobility (the numbers are presented as standardised mean difference and 95% confidence interval)

Cons		-0.32 (-0.92,0.29)	
-0.06 (-0.84,0.72)	CD+Fu	-0.26 (-0.75,0.24)	
-0.32 (-0.92,0.29)	-0.26 (-0.75,0.24)	CD	-0.18 (-0.61,0.26)
-0.49 (-1.24,0.25)	-0.43 (-1.08,0.21)	-0.18 (-0.61,0.26)	MPD

The results from bottom-left are network evidence. The results from top-right are pairwise evidence. For the network evidence, the minus sign means the up to the left intervention has the superior result. For the pairwise evidence, the minus sign means the leftward intervention has the superior result.

Adverse effect due to any reason (the numbers are presented as odds ratio and 95% confidence interval)

Cons			0.68 (0.37,1.25)					0.58 (0.22,1.51)
0.38 (0.10,1.45)	CD+ID		1.07 (0.41,2.77)					
0.44 (0.12,1.64)	1.16 (0.29,4.62)	MSD	0.84 (0.29,2.4)	5 (0.89, 28.07)				
0.41 (0.16,1.05)	1.07 (0.41,2.77)	0.92 (0.34,2.50)	CD	**3.78 (1.69,8.45)**			0.84 (0.39, 1.85)	1.6 (0.44, 5.83)
1.21 (0.41,3.54)	3.18 (0.97,10.49)	2.73 (0.88,8.52)	**2.98 (1.45,6.12)**	MPD		1.91 (0.75,4.88)	1.48 (0.08, 27.94)	2.05 (0.43,9.86)
1.57 (0.14,18.16)	4.14 (0.35,48.54)	3.55 (0.31,41.06)	3.87 (0.40,37.50)	1.30 (0.15,11.37)	EpiS	1.47 (0.21,10.38)		
2.31 (0.53,10.06)	**6.08 (1.36,27.18)**	**5.22 (1.20,22.76)**	**5.68 (1.79,18.09)**	1.91 (0.75,4.88)	1.47 (0.21,10.38)	Endo		
0.31 (0.10,0.94)	0.82 (0.25,2.70)	0.70 (0.21,2.36)	0.77 (0.37,1.57)	0.26 (0.10,0.67)	0.20 (0.02,2.11)	0.13 (0.04,0.51)	CD+Fu	**7.85 (1.77,34.80)**
1.00 (0.41,2.42)	2.62 (0.75,9.20)	2.25 (0.65,7.79)	**2.45 (1.08,5.56)**	0.82 (0.32,2.10)	0.63 (0.06,6.80)	0.43 (0.11,1.65)	**3.20 (1.25,8.22)**	ID

The results from bottom-left are network evidence. The results from top-right are pairwise evidence. For the network evidence, the value less than one means the up to the left intervention has the superior result. For the pairwise evidence, the minus sign means the leftward intervention has the superior result. The statistically significant results were presented in bold.

Intervention related adverse effect (the numbers are presented as odds ratio and 95% confidence interval)

Cons							**0.04 (0.002, 0.68)**	0.2 (0.03, 1.29)
0.06 (0.01,0.33)	CD+ID		1 (0.53,1.87)				0.66 (0.37,1.16)	
0.06 (0.01,0.39)	0.99 (0.35,2.80)	MSD	0.8 (0.3,2.14)	**5.58 (1.19,26.27)**				
0.05 (0.01,0.27)	0.88 (0.55,1.41)	0.89 (0.35,2.25)	CD	**3.32 (1.71,6.46)**			0.94 (0.6, 1.48)	1.52 (0.52,4.47)
0.15 (0.03,0.85)	**2.55 (1.20,5.44)**	2.58 (0.91,7.33)	**2.91 (1.59,5.32)**	MPD		1.78 (0.66,4.80)	1.48 (0.09, 25.33)	2.05 (0.48, 8.76)
0.19 (0.01,2.70)	3.09 (0.35,27.71)	3.13 (0.31,31.43)	3.52 (0.41,30.10)	1.21 (0.15,9.48)	EpiS	1.47 (0.24,8.93)		
0.27 (0.04,1.97)	**4.55 (1.31,15.82)**	**4.60 (1.09,19.39)**	**5.18 (1.62,16.54)**	1.78 (0.66,4.80)	1.47 (0.24,8.93)	Endo		
0.04 (0.01,0.24)	0.74 (0.47,1.17)	0.75 (0.27,2.05)	0.84 (0.57,1.25)	**0.29 (0.14,0.58)**	0.24 (0.03,2.11)	**0.16 (0.05,0.55)**	CD+Fu	**7.85 (1.95, 31.55)**
0.18 (0.04,0.84)	**2.93 (1.25,6.87)**	2.96 (0.91,9.58)	**3.33 (1.58,7.02)**	1.15 (0.49,2.69)	0.95 (0.10,8.79)	0.64 (0.17,2.38)	**3.95 (1.80,8.67)**	ID

The results from bottom-left are network evidence. The results from top-right are pairwise evidence. For the network evidence, the value less than one means the up to the left intervention has the superior result. For the pairwise evidence, the minus sign means the leftward intervention has the superior result. The statistically significant results were presented in bold.

Reoperation rate (the numbers are presented as odds ratio and 95% confidence interval)

CD		1.05 (0.32,3.48)	2.7 (0.76,10)	0.97 (0.54,1.76)	**0.2 (0.09, 0.43)**

1.64 (0.30,8.95)	Endo		1.31 (0.34,5.01)		
0.92 (0.29,2.95)	0.56 (0.07,4.21)	MSD	5.64 (0.46,69.3)		
2.16 (0.76,6.12)	1.31 (0.34,5.01)	2.34 (0.52,10.55)	MPD		**0.15 (0.03,0.75)**
0.97 (0.54,1.76)	0.59 (0.10,3.58)	1.06 (0.29,3.91)	0.45 (0.14,1.50)	CD+Fu	
0.22 (0.10,0.45)	**0.13 (0.02,0.75)**	**0.24 (0.06,0.92)**	**0.10 (0.03,0.30)**	**0.22 (0.09,0.57)**	ID

The results from bottom-left are network evidence. The results from top-right are pairwise evidence. For the network evidence, the value less than one means the up to the left intervention has the superior result. For the pairwise evidence, the minus sign means the leftward intervention has the superior result. The statistically significant results were presented in bold.

Treatment withdrawal due to any reason (the numbers are presented as odds ratio and 95% confidence interval)

Cons			1.64 (0.59,4.54)	0.63 (0.35, 1.14)				1.05 (0.68,1.62)
0.92 (0.17,4.98)	MSD							1.19 (0.23,6.12)
0.96 (0.39,2.36)	1.04 (0.16,6.88)	CD+ID		0.70 (0.33,1.48)				
1.09 (0.53,2.25)	1.18 (0.20,7.09)	1.14 (0.37,3.54)	ID		0.82 (0.2,3.29)			2.72 (0.82,8.99)
0.67 (0.40,1.12)	0.73 (0.13,4.13)	0.70 (0.33,1.48)	0.62 (0.26,1.45)	CD+Fu				1.42 (0.57,3.51)
1.52 (0.73,3.18)	1.65 (0.28,9.64)	1.59 (0.51,4.91)	1.40 (0.60,3.26)	2.26 (0.97,5.29)	MPD		1.67 (0.77,3.62)	0.62 (0.29, 1.31)
0.85 (0.12,5.87)	0.92 (0.07,11.34)	0.88 (0.11,7.34)	0.78 (0.11,5.64)	1.26 (0.17,9.14)	0.56 (0.09,3.34)	EpiS	3.00 (0.60,15.11)	

2.54 (0.87,7.38)	2.75 (0.40,18.89)	2.65 (0.67,10.41)	2.33 (0.74,7.34)	**3.77 (1.20,11.90)**	1.67 (0.77,3.62)	3.00 (0.60,15.11)	Endo	
1.09 (0.74,1.61)	1.19 (0.23,6.12)	1.14 (0.45,2.91)	1.00 (0.49,2.05)	1.63 (0.93,2.83)	0.72 (0.38,1.37)	1.29 (0.19,8.70)	0.43 (0.16,1.18)	CD

The results from bottom-left are network evidence. The results from top-right are pairwise evidence. For the network evidence, the value less than one means the up to the left intervention has the superior result. For the pairwise evidence, the minus sign means the leftward intervention has the superior result.

Appendix S12: Inconsistency test

Short-term physical function

```
Testing for inconsistency:
 ( 1)  [_y_06]des_050609 = 0
 ( 2)  [_y_13]des_0613 = 0
 ( 3)  [_y_04]des_0413 = 0
 ( 4)  [_y_10]des_0810 = 0
 ( 5)  [_y_04]des_0304 = 0

           chi2(  5) =   88.11
         Prob > chi2 =    0.0000
```

Side	Direct		Indirect		Difference			tau
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	P>|z|	
01 02 *	.2143959	.6765654	-1.288517	631.7054	1.502913	631.7058	0.998	.6660866
02 05 *	.3163835	.7145477	-.4596292	283.043	.7760127	283.0439	0.998	.6660866
03 04	.5436922	.4950356	1.627508	.8494022	-1.083815	.982908	0.270	.6576354
03 06	-3.78e-08	.4787567	-1.083808	.8584238	1.083808	.9829036	0.270	.6576338
04 08	.490507	.7689217	-.6518365	1.046625	1.142344	1.298717	0.379	.6694671
04 13	-2.46343	.6281749	-.6536931	.6669049	-1.809737	.9161692	0.048	.5977653
05 11	.184862	.3979227	1.32695	1.2361	-1.142088	1.298601	0.379	.6694622
05 06 *	-.4488315	.2793662	1.475571	.5346112	-1.924402	.6022733	0.001	.4865108
05 09 *	.1642495	.7271731	.514558	1.46312	-.3503085	1.633607	0.830	.6918545
05 13	.30807	.2661972	-2.119712	.3748986	2.427782	.4596367	0.000	.3291609
06 09 *	.3695614	.7275027	.0192556	1.462629	.3503058	1.633608	0.830	.6918544
06 13	-1.54675	.6575218	-.0052125	.4735583	-1.541538	.8103039	0.057	.6060174
07 08 *	.1470555	.3500384	1.613997	316.5224	-1.466942	316.5226	0.996	.6660867
08 11	-.3995542	.7771596	-1.541802	1.040492	1.142248	1.298692	0.379	.6694663
08 10	-.5123602	.6512038	-2.876769	.9574953	2.364409	1.157957	0.041	.5976542
08 12	-2.147218	.6986632	.2171976	.9234349	-2.364415	1.157956	0.041	.5976534
10 12	.7295672	.6547312	-1.634856	.9550889	2.364423	1.157958	0.041	.5976534

```
* Warning: all the evidence about these contrasts comes from the trials which directly compare th
See help file for more information.
```

Long-term physical function

```
Testing for inconsistency:
 ( 1)  [_y_3]des_37 = 0
 ( 2)  [_y_7]des_67 = 0
 ( 3)  [_y_7]des_678 = 0
 ( 4)  [_y_8]des_68 = 0
 ( 5)  [_y_8]des_78 = 0
 ( 6)  [_y_3]des_13 = 0
 ( 7)  [_y_2]des_25 = 0
 ( 8)  [_y_4]des_34 = 0

           chi2(  8) =   17.26
         Prob > chi2 =    0.0275
```

Side		Direct Coef.	Std. Err.	Indirect Coef.	Std. Err.	Difference Coef.	Std. Err.	P>|z|	tau
1	2	1.740321	.4376922	1.056283	.8113991	.6840377	.9219583	0.458	.577883
1	3	.1718247	.6397699	.5347258	.4678497	-.3629011	.7925837	0.647	.5851253
1	6	-.2687397	.3661587	1.131587	.3900463	-1.400327	.5350284	0.009	.4841182
1	7	1.793402	.5329677	-.0872644	.3196124	1.880667	.6214553	0.002	.4621412
2	5	-.7154008	.5913133	-1.399438	.7073599	.6840374	.9219595	0.458	.5778831
3	4	-.1540932	.596434	.3287703	.5520548	-.4828635	.8127103	0.552	.5845504
3	6	-.0165286	.4406934	-.0037099	.427671	-.0128187	.6140573	0.983	.5891204
3	7	.0843233	.662585	-.0272614	.4032553	.1115848	.7756505	0.886	.5872914
4	6	-.257114	.4389557	.2257458	.6838339	-.4828598	.8127092	0.552	.5845503
5	7	-.0331745	.4244176	-.7172092	.8184587	.6840346	.9219576	0.458	.5778831
6	9 *	.2932904	.3426354	.0162909	64.69054	.2769995	64.69145	0.997	.5701316
6	7	-.3391365	.2493008	.7670637	.3579503	-1.1062	.4360185	0.011	.4935416
6	8	-.2188091	.3253949	-.3448924	.6518089	.1260833	.7285384	0.863	.5880867
7	8	-.0971996	.4476453	-.4253374	.4592397	.3281378	.6415505	0.609	.5806583

* Warning: all the evidence about these contrasts comes from the trials which directly compare t
> hem.

All-cause mortality

```
Testing for inconsistency:
 ( 1)  [_y_4]des_145 = 0
 ( 2)  [_y_3]des_23 = 0

           chi2(  2) =    0.16
         Prob > chi2 =    0.9213
```

Side		Direct Coef.	Std. Err.	Indirect Coef.	Std. Err.	Difference Coef.	Std. Err.	P>|z|	tau
2	3	-.9699002	.6998908	-1.525724	1.180772	.5558241	1.372614	0.686	2.70e-07
1	2	-.2011717	.4483543	-.7569936	1.297322	.5558219	1.372613	0.686	8.05e-08
1	3	-1.726896	1.092338	-1.171072	.8311851	-.5558242	1.372614	0.686	1.72e-07
1	4 *	1.129607	1.168047	-.9364888	2696.223	2.066096	2696.224	0.999	2.58e-08
1	5 *	1.123614	1.648234	1.14769	4.05557	-.0240764	4.672231	0.996	1.47e-07
1	6 *	1.151958	1.638431	-.8142662	6334.693	1.966224	6334.694	1.000	1.32e-06
4	5 *	8.56e-11	1.176421	-.0240798	4.521701	.0240798	4.672232	0.996	2.07e-07

* Warning: all the evidence about these contrasts comes from the trials which directly compare them.
See help file for more information.

Short-term back pain

```
Testing for inconsistency:
 ( 1)  [_y_2]des_123 = 0
 ( 2)  [_y_3]des_23 = 0
 ( 3)  [_y_4]des_24 = 0

           chi2(  3) =  426.89
         Prob > chi2 =    0.0000
```

Side	Direct		Indirect		Difference			tau
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	P>|z|	
1 7 *	2.5	1.736782	-.3698021	92.05437	2.869802	92.07076	0.975	1.630408
1 2	-.997353	.9128804	1.607266	1.931702	-2.604619	2.136642	0.223	1.525742
1 3	-.1035601	1.873558	-.7275673	2.172894	.6240072	2.86869	0.828	1.865874
1 4	.1000001	1.874349	-1.0102	2.117138	1.1102	2.827624	0.695	1.831434
1 5 *	.7999998	1.804214	2.760746	760.5333	-1.960746	760.5355	0.998	1.630321
2 3 *	.4466082	1.209982	-2.695297	3.710551	3.141905	3.900878	0.421	1.704188
2 4	-.3399999	1.833938	.7709022	2.152339	-1.110902	2.827701	0.694	1.831447
5 6 *	.5999999	1.660577	3.601677	1661.065	-3.001677	1661.066	0.999	1.63032

* Warning: all the evidence about these contrasts comes from the trials which directly compare t
> hem.

Long-term back pain

```
Testing for inconsistency:
 ( 1)  [_y_1]des_15 = 0
 ( 2)  [_y_2]des_25 = 0
 ( 3)  [_y_3]des_35 = 0
 ( 4)  [_y_4]des_45 = 0
 ( 5)  [_y_2]des_12 = 0

           chi2(  5) =   15.99
         Prob > chi2 =    0.0069
```

Side	Direct		Indirect		Difference			tau
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	P>|z|	
1 2	-.2180909	.875505	1.606518	1.469568	-1.824609	1.71055	0.286	1.232046
1 5	.4414401	.9781778	.7744236	1.544105	-.3329835	1.828117	0.855	1.343286
2 4	-.8500001	1.285526	.5097397	1.128239	-1.35974	1.71041	0.427	1.28205
2 5	.6024718	.7518373	-.5572722	1.186229	1.159744	1.404662	0.409	1.266676
3 4	-.8000002	1.52692	-.7385815	1.6288	-.0614187	2.232594	0.978	1.335311
3 5	-.3999998	1.391409	-.4614468	1.745979	.061447	2.232591	0.978	1.335311
3 6 *	-.0999999	1.277777	3.790601	1589.956	-3.890601	1589.957	0.998	1.243962
4 5	-.1078245	.9850743	1.017486	1.186424	-1.12531	1.541769	0.465	1.287215
5 7 *	2.3	1.385174	-.2747673	63.28752	2.574767	63.30267	0.968	1.244014

* Warning: all the evidence about these contrasts comes from the trials which directly compare t
> hem.

Short-term leg pain

```
Testing for inconsistency:
 ( 1)  [_y_2]des_124 = 0
 ( 2)  [_y_3]des_23 = 0

           chi2(  2) =   18.96
         Prob > chi2 =    0.0001
```

Side	Direct		Indirect		Difference			tau
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	P>|z|	
1 5 *	1.8	.8642363	-.1188441	92.85298	1.918844	92.857	0.984	.5562167
1 2 *	-.2866378	.1013921	1.388496	.3753993	-1.675134	.3888508	0.000	1.02e-06
1 3	.4284838	.363034	-1.246635	.1393291	1.675118	.3888525	0.000	1.60e-07
1 4 *	-.8700002	.6040982	.3464695	1.369863	-1.21647	1.498833	0.417	.5966026
2 3	-.96	.0955628	.7151263	.3769269	-1.675126	.3888523	0.000	8.00e-08
2 4 *	-.5500002	.6053654	-1.766376	1.368197	1.216376	1.498843	0.417	.5966048
2 6 *	-.3001058	1.486771	4.189369	6780.698	-4.489475	6780.699	0.999	.5562141

* Warning: all the evidence about these contrasts comes from the trials which directly compare t
> hem.

Long-term leg pain

```
Testing for inconsistency:
 ( 1)  [_y_1]des_15 = 0
 ( 2)  [_y_2]des_25 = 0
 ( 3)  [_y_3]des_35 = 0
 ( 4)  [_y_4]des_45 = 0
 ( 5)  [_y_2]des_12 = 0

           chi2(  5) =   10.84
         Prob > chi2 =    0.0547
```

Side	Direct		Indirect		Difference			tau
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	P>|z|	
1 2	.0245783	.2861463	2.448357	.9010492	-2.423779	.9432629	0.010	.3985835
1 5	.7546885	.6514464	-.3760377	.7900338	1.130726	1.016605	0.266	.6739179
2 4	-.3400005	.7473858	-.0451218	.8202223	-.2948786	1.109662	0.790	.741391
2 5	.0469761	.4385107	-.3166969	.8693022	.3636731	.9674601	0.707	.7321612
2 8 *	.0517736	1.505726	3.958248	5552.593	-3.906474	5552.594	0.999	.6765376
3 4	-1	2.050408	-.2558617	1.104851	-.7441384	2.329135	0.749	.6983202
3 5	-.0999998	.9589591	-.8441572	2.122559	.7441574	2.329132	0.749	.6983295
3 6 *	-.03	.7383375	2.976293	1620.494	-3.006293	1620.494	0.999	.6765377
4 5	-.0257494	.7411706	.4233545	.7959963	-.4491039	1.082586	0.678	.7338075
5 7 *	1.7	.9318363	-.1109572	52.52371	1.810957	52.53197	0.972	.6765426

* Warning: all the evidence about these contrasts comes from the trials which directly compare t
> hem.

Short-term overall pain

Not available

Short-term global impression of recovery

```
Testing for inconsistency:
 ( 1)  [_y_4]des_345 = 0

           chi2(  1) =    0.96
         Prob > chi2 =    0.3269
```

Side		Direct Coef.	Direct Std. Err.	Indirect Coef.	Indirect Std. Err.	Difference Coef.	Difference Std. Err.	P>|z|
1	2 *	-1.280105	.4180844	-4.641748	927.4684	3.361644	927.4689	0.997
1	4 *	-.6895174	.5901875	1.051039	437.7745	-1.740556	437.7749	0.997
3	6 *	-.7230001	.4076242	.7581038	336.2675	-1.481104	336.2677	0.996
3	4 *	.8215773	.4574267	-.9107226	391.6779	1.7323	391.6788	0.996
3	5 *	.1381503	.5878441	-1.693565	1.779264	1.831716	1.868494	0.327
4	5 *	-1.275543	.7566978	.5561718	1.576302	-1.831715	1.868503	0.327

* Warning: all the evidence about these contrasts comes from the trials which directly compare them.

Long-term global impression of recovery

```
Testing for inconsistency:
 ( 1)  [_y_1]des_124 = 0
 ( 2)  [_y_4]des_24 = 0

           chi2(  2) =    6.65
         Prob > chi2 =    0.0360
```

Side		Direct Coef.	Direct Std. Err.	Indirect Coef.	Indirect Std. Err.	Difference Coef.	Difference Std. Err.	P>|z|
5	6	.	.	.	.	.	.	.
1	2 *	-1.137183	.6132607	2.857318	2.60181	-3.994501	2.664781	0.134
1	4	2.277707	.9264561	-1.101271	.7753898	3.378977	1.310385	0.010
2	5 *	-.6915786	.8332514	.4599117	334.4181	-1.15149	334.4191	0.997
2	3 *	-.0870114	.821581	-1.389248	1277.837	1.302236	1277.837	0.999
2	4 *	1.203802	.3677379	4.908388	1.775285	-3.704585	1.753663	0.035

* Warning: all the evidence about these contrasts comes from the trials which directly compare t
> hem.

Short-term mobility

Not available

Long-term mobility

Not available

Adverse effect due to any reason

```
Testing for inconsistency:
 ( 1)  [_y_7]des_27 = 0
 ( 2)  [_y_6]des_36 = 0
 ( 3)  [_y_7]des_37 = 0
 ( 4)  [_y_7]des_67 = 0
 ( 5)  [_y_7]des_678 = 0
 ( 6)  [_y_8]des_78 = 0

           chi2(  6) =    9.73
         Prob > chi2 =    0.1365
```

Side	Direct		Indirect		Difference			tau
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	P>|z|	
2 7	.3831892	.3114058	1.719776	.6259748	-1.336587	.6991552	0.056	8.11e-07
1 2	-.5463807	.4875635	.7902156	.5011013	-1.336596	.6991572	0.056	3.49e-07
1 3	2.063693	.7563752	.667909	.5299411	1.395784	.9235481	0.131	.2603931
1 6	.7178398	.8017699	-.6390385	.5477878	1.356878	.9710337	0.162	.3056873
1 7	.4656006	.6616624	1.22488	.5686488	-.7592794	.8597824	0.377	.4552634
3 6	.3856625	1.502884	-1.565809	.5301946	1.951472	1.593664	0.221	.3997943
3 7	-.1666599	.4044178	-.6590073	.8622298	.4923474	.9402446	0.601	.3747952
4 5 *	.3852624	.9974396	3.57754	5434.449	-3.192278	5434.449	1.000	.3848156
4 6 *	.6472435	.4788679	-.2619777	1098.769	.9092213	1098.769	0.999	.3848602
6 7	1.328921	.4083769	.1465266	.7476921	1.182394	.8288693	0.154	.3508373
6 8	1.614289	.8779043	.4440717	.8423495	1.170218	1.274044	0.358	.3752703
7 8 *	-.1749926	.5359633	.95613	1.890953	-1.131123	1.963696	0.565	.4028228
7 9 *	.067372	.4845525	-1.788221	1417.081	1.855593	1417.081	0.999	.3848602

```
* Warning: all the evidence about these contrasts comes from the trials which directly compare t
> hem.
```

Intervention related adverse effect

```
Testing for inconsistency:
 ( 1)  [_y_7]des_27 = 0
 ( 2)  [_y_6]des_36 = 0
 ( 3)  [_y_7]des_37 = 0
 ( 4)  [_y_7]des_67 = 0
 ( 5)  [_y_7]des_678 = 0
 ( 6)  [_y_8]des_78 = 0
 ( 7)  [_y_9]des_79 = 0

           chi2(  7) =    7.78
         Prob > chi2 =    0.3524
```

Side		Direct		Indirect		Difference			tau
		Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	P>|z|	
2	7	3.254683	1.461077	2.793185	1.02315	.4614977	1.783698	0.796	3.61e-06
1	2	-1.611125	.9462998	-2.072437	1.511959	.4613117	1.783677	0.796	3.26e-07
1	3	2.063693	.7101401	1.050735	.4857099	1.012958	.8603563	0.239	8.25e-07
1	6	.7178397	.7412086	-.1730058	.5398422	.8908454	.9169623	0.331	5.67e-08
1	7	.4165669	.5532316	1.905356	.5220747	-1.488789	.7606755	0.050	1.88e-07
3	6	.3856625	1.448732	-1.342469	.3678301	1.728131	1.494698	0.248	2.31e-07
3	7	-.0620679	.2347828	-.4407156	.3730095	.3786477	.4407483	0.390	3.59e-06
3	9	-.4160235	.2925055	-.0949739	.3892409	-.3210496	.4868963	0.510	3.66e-06
4	5 *	.3852624	.9202188	-2.571512	5369.119	2.956774	5369.119	1.000	8.73e-07
4	6 *	.5764266	.5060415	3.292269	1265.438	-2.715842	1265.438	0.998	2.77e-06
6	7	1.198642	.3384288	.4794483	.70514	.7191937	.7747676	0.353	7.57e-08
6	8	1.723661	.7916132	.0213549	.8812684	1.702306	1.288875	0.187	4.42e-08
7	8 *	-.219571	.4954795	1.007063	1.695749	-1.226634	1.773773	0.489	8.97e-09
7	9	.0067644	.3185313	-.3142864	.3682469	.3210508	.4868963	0.510	2.33e-07

* Warning: all the evidence about these contrasts comes from the trials which directly compare them.

Reoperation rate

```
Testing for inconsistency:
 ( 1)  [_y_4]des_24 = 0
 ( 2)  [_y_4]des_245 = 0
 ( 3)  [_y_5]des_25 = 0

           chi2(  3) =    1.45
         Prob > chi2 =    0.6950
```

Side		Direct		Indirect		Difference			tau
		Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	P>|z|	
2	3	.	.	.	.	.	.	.	.
2	4	-.9883957	.6505924	-.3730566	.8560406	-.6153391	1.044439	0.556	2.17e-08
2	5 *	-.0454539	.605347	3.51678	3.145716	-3.562234	3.203431	0.266	1.33e-06
1	2	-1.630604	.3976699	-.8134771	1.043022	-.8171266	1.11626	0.464	3.20e-08
1	4	-1.867267	.8075512	-2.684393	.7706481	.8171261	1.11626	0.464	4.55e-06
4	5	1.734774	1.280994	.0358334	1.216155	1.698941	1.967642	0.388	3.97e-07
4	6 *	.2736965	.6828045	1.532376	2855.699	-1.25868	2855.699	1.000	1.62e-06

* Warning: all the evidence about these contrasts comes from the trials which directly compare t
> hem.

Treatment withdrawal due to any reason

```
Testing for inconsistency:
 ( 1)  [_y_6]des_26 = 0
 ( 2)  [_y_7]des_27 = 0
 ( 3)  [_y_7]des_57 = 0

           chi2(  3) =    4.07
         Prob > chi2 =    0.2545
```

Side	Direct		Indirect		Difference			tau
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	P>|z|	
2 6	.4562084	.3003674	.2295237	.508869	.2266847	.5909046	0.701	1.19e-06
2 7	-.4948253	.5157087	.3420303	.5280657	-.8368556	.7381117	0.257	2.16e-07
1 2	.0452007	.2229226	.2571691	.4351252	-.2119684	.4889053	0.665	.0000478
1 5	-.4700493	.3618037	.3823411	.8180639	-.8523904	.8945001	0.341	2.43e-07
1 6	.3481421	.4623081	.5748267	.3680211	-.2266846	.5909047	0.701	4.25e-08
1 7	.9852836	.6145072	-.5360393	.4551899	1.521323	.7647332	0.047	1.13e-07
1 9 *	.169899	.8374896	.1616159	1661.089	.0082831	1661.089	1.000	1.10e-09
3 4 *	1.098612	.8247861	1.568877	4273.35	-.4702648	4273.35	1.000	4.00e-07
3 5 *	.5106256	.3953026	-.4456258	1206.053	.9562514	1206.053	0.999	7.93e-07
5 7	-.1984509	.7066755	.6539395	.548399	-.8523904	.8945008	0.341	8.86e-08
6 8 *	-.3553303	.3811758	-.7465917	2086.813	.3912614	2086.813	1.000	1.79e-06

* Warning: all the evidence about these contrasts comes from the trials which directly compare them.

Appendix S13. Results for sensitivity analyses

Short-term physical function (the numbers are presented as standardised mean difference and 95% confidence interval)

1.Exclude studies received commercial funding (Benyamin 2016, Manchikanti 2009, Manchikanti 2015, Moojen 2015 and Zucherman 2005)

Cons												
0.92 (-0.05,1.90)	ID											
-0.62 (-2.51,1.27)	-1.55 (-3.60,0.51)	EpiASH										
0.11 (-1.42,1.64)	-0.82 (-2.55,0.92)	0.73 (-0.38,1.84)	BT									
0.01 (-1.10,1.13)	-0.91 (-2.11,0.29)	0.63 (-1.49,2.76)	-0.10 (-1.91,1.72)	MSD								
-0.40 (-1.46,0.65)	-1.33 (-2.67,0.01)	0.22 (-1.34,1.78)	-0.51 (-1.61,0.59)	-0.42 (-1.86,1.03)	EpiAS							
-0.22 (-1.45,1.00)	-1.15 (-2.62,0.33)	0.40 (-1.28,2.08)	-0.33 (-1.59,0.93)	-0.23 (-1.80,1.34)	0.18 (-0.44,0.80)	EpiA						
0.38 (-0.37,1.12)	-0.55 (-1.35,0.26)	1.00 (-0.94,2.94)	0.27 (-1.33,1.86)	0.36 (-0.60,1.33)	0.78 (-0.37,1.93)	0.60 (-0.71,1.90)	MPD					

0.18 (-0.39,0.76)	-0.74 (-1.54,0.06)	0.80 (-1.12,2.72)	0.07 (-1.50,1.64)	0.17 (-0.79,1.13)	0.59 (-0.53,1.71)	0.40 (-0.87,1.68)	-0.19 (-0.71,0.32)	CD				
0.08 (-1.09,1.26)	-0.84 (-2.18,0.50)	0.70 (-1.18,2.58)	-0.03 (-1.54,1.49)	0.07 (-1.37,1.51)	0.49 (-0.56,1.53)	0.31 (-0.91,1.52)	-0.29 (-1.41,0.83)	-0.10 (-1.25,1.05)	EpiS			
0.38 (-0.58,1.33)	-0.55 (-1.60,0.50)	1.00 (-0.96,2.96)	0.27 (-1.35,1.88)	0.37 (-0.81,1.54)	0.78 (-0.40,1.96)	0.60 (-0.73,1.93)	0.00 (-0.69,0.69)	0.20 (-0.63,1.02)	0.29 (-0.72,1.31)	Endo		
0.50 (-0.74,1.74)	-0.43 (-1.79,0.93)	1.12 (-1.10,3.33)	0.39 (-1.53,2.31)	0.49 (-0.98,1.95)	0.90 (-0.67,2.47)	0.72 (-0.97,2.41)	0.12 (-1.09,1.34)	0.32 (-0.78,1.42)	0.42 (-1.18,2.01)	0.12 (-1.26,1.50)	CD+Fu	
0.71 (-0.88,2.31)	-0.21 (-1.90,1.48)	1.33 (-1.10,3.76)	0.60 (-1.56,2.77)	0.70 (-1.07,2.47)	1.12 (-0.75,2.98)	0.94 (-1.03,2.90)	0.34 (-1.24,1.91)	0.53 (-0.96,2.02)	0.63 (-1.25,2.51)	0.34 (-1.37,2.04)	0.21 (-0.79,1.22)	CD+ID

2.Exclude studies with level 1 only (Yagi 2009)

Cons												
0.79 (-0.15,1.73)	ID											
0.40 (-1.18,1.97)	-0.39 (-2.05,1.27)	EpiASH										
0.37 (-1.18,1.92)	-0.42 (-2.05,1.22)	-0.03 (-1.09,1.04)	BT									
-0.07 (-1.38,1.24)	-0.86 (-2.13,0.42)	-0.47 (-2.40,1.47)	-0.44 (-2.35,1.47)	MSD								
-0.89 (-2.02,0.24)	**-1.68 (-2.92,-0.43)**	**-1.29 (-2.39,-0.19)**	**-1.26 (-2.32,-0.20)**	-0.82 (-2.41,0.77)	EpiAS							
-0.74 (-2.04,0.55)	**-1.53 (-2.93,-0.14)**	-1.14 (-2.41,0.12)	-1.12 (-2.35,0.12)	-0.68 (-2.38,1.03)	0.15 (-0.48,0.77)	EpiA						
0.09 (-0.81,0.99)	-0.70 (-1.46,0.07)	-0.31 (-1.95,1.34)	-0.28 (-1.90,1.34)	0.16 (-0.98,1.29)	0.98 (-0.24,2.21)	0.84 (-0.54,2.21)	MPD					
0.31 (-0.37,0.98)	-0.48 (-1.19,0.22)	-0.09 (-1.70,1.51)	-0.06 (-1.64,1.52)	0.37 (-0.76,1.51)	1.20 (0.03,2.37)	1.05 (-0.27,2.38)	0.22 (-0.42,0.85)	CD				

-0.88 (-1.94,0.19)	**-1.67 (-2.57,-0.76)**	-1.28 (-2.84,0.29)	-1.25 (-2.79,0.30)	-0.81 (-2.20,0.58)	0.01 (-1.11,1.13)	-0.13 (-1.42,1.15)	**-0.97 (-1.87,-0.07)**	**-1.18 (-2.12,-0.25)**	EpiS			
-0.11 (-1.18,0.96)	-0.90 (-1.84,0.04)	-0.51 (-2.18,1.16)	-0.48 (-2.12,1.16)	-0.04 (-1.38,1.29)	0.78 (-0.48,2.03)	0.63 (-0.77,2.03)	-0.20 (-0.96,0.55)	-0.42 (-1.32,0.48)	0.77 (-0.01,1.54)	Endo		
0.62 (-0.83,2.07)	-0.17 (-1.63,1.30)	0.23 (-1.83,2.28)	0.25 (-1.78,2.29)	0.69 (-1.02,2.40)	1.51 (-0.22,3.25)	1.37 (-0.48,3.21)	0.53 (-0.90,1.96)	0.32 (-0.97,1.60)	1.50 (-0.09,3.09)	0.74 (-0.83,2.30)	CD+Fu	
0.84 (-1.05,2.72)	0.05 (-1.85,1.94)	0.44 (-1.94,2.82)	0.47 (-1.90,2.83)	0.91 (-1.19,3.00)	1.73 (-0.38,3.84)	1.58 (-0.62,3.78)	0.75 (-1.12,2.62)	0.53 (-1.23,2.29)	1.72 (-0.28,3.71)	0.95 (-1.03,2.93)	0.21 (-0.99,1.42)	CD+ID

3.Exclude studies with mixed type of disease (Benyamin 2016, Song 2016, Weinstein 2008)

Cons												
0.78 (-0.57,2.13)	ID											
0.91 (-1.16,2.98)	0.13 (-2.34,2.60)	EpiASH										
0.87 (-1.18,2.92)	0.10 (-2.36,2.55)	-0.04 (-1.29,1.22)	BT									
0.01 (-1.69,1.71)	-0.77 (-2.30,0.76)	-0.90 (-3.58,1.78)	-0.86 (-3.53,1.80)	MSD								
-0.40 (-2.03,1.23)	-1.18 (-3.29,0.94)	**-1.31 (-2.59,-0.02)**	**-1.27 (-2.52,-0.02)**	-0.41 (-2.76,1.95)	EpiAS							
-0.25 (-2.04,1.54)	-1.03 (-3.27,1.21)	-1.16 (-2.64,0.33)	-1.12 (-2.58,0.33)	-0.26 (-2.73,2.21)	0.15 (-0.60,0.89)	EpiA						
0.29 (-0.97,1.54)	-0.49 (-1.40,0.42)	-0.62 (-3.04,1.80)	-0.58 (-2.99,1.82)	0.28 (-1.06,1.62)	0.69 (-1.37,2.74)	0.54 (-1.65,2.72)	MPD					
0.26 (-0.78,1.30)	-0.52 (-1.37,0.34)	-0.65 (-2.97,1.67)	-0.61 (-2.91,1.69)	0.25 (-1.09,1.60)	0.66 (-1.27,2.59)	0.51 (-1.56,2.58)	-0.03 (-0.72,0.67)	CD				
-1.05 (-2.74,0.65)	**-1.82 (-3.04,-0.61)**	-1.95 (-4.63,0.73)	-1.92 (-4.58,0.75)	-1.05 (-2.87,0.76)	-0.65 (-3.00,1.71)	-0.79 (-3.26,1.67)	**-1.33 (-2.63,-0.04)**	-1.31 (-2.65,0.04)	EpiS			
-0.03 (-1.56,1.51)	-0.80 (-2.00,0.39)	-0.93 (-3.51,1.65)	-0.90 (-3.46,1.67)	-0.03 (-1.66,1.59)	0.37 (-1.87,2.61)	0.22 (-2.14,2.58)	-0.31 (-1.26,0.64)	-0.29 (-1.42,0.84)	1.02 (-0.23,2.27)	Endo		
0.58 (-1.27,2.42)	-0.20 (-1.94,1.54)	-0.33 (-3.10,2.44)	-0.30 (-3.05,2.46)	0.57 (-1.46,2.60)	0.98 (-1.48,3.43)	0.83 (-1.74,3.40)	0.29 (-1.38,1.96)	0.32 (-1.20,1.84)	1.62 (-0.40,3.65)	0.60 (-1.29,2.50)	CD+Fu	
0.79 (-1.55,3.14)	0.01 (-2.25,2.28)	-0.12 (-3.25,3.01)	-0.08 (-3.20,3.04)	0.78 (-1.71,3.28)	1.19 (-1.66,4.04)	1.04 (-1.91,3.99)	0.50 (-1.71,2.71)	0.53 (-1.57,2.63)	1.84 (-0.66,4.33)	0.82 (-1.57,3.20)	0.21 (-1.24,1.66)	CD+ID

4. Exclude studies which has medication therapy in the conservative care group (Slatis 2011 and Weinstein 2008)

Cons												
0.96 (-0.47,2.40)	ID											
0.53 (-1.32,2.39)	-0.43 (-2.42,1.56)	EpiASH										
0.50 (-1.34,2.33)	-0.47 (-2.44,1.50)	-0.03 (-1.28,1.21)	BT									
0.13 (-1.68,1.94)	-0.83 (-2.33,0.67)	-0.40 (-2.71,1.92)	-0.36 (-2.66,1.93)	MSD								
-0.77 (-2.12,0.58)	**-1.74 (-3.27,-0.20)**	**-1.31 (-2.58,-0.04)**	**-1.27 (-2.51,-0.03)**	-0.91 (-2.84,1.03)	EpiAS							
-0.63 (-2.17,0.91)	-1.59 (-3.29,0.11)	-1.16 (-2.63,0.31)	-1.12 (-2.56,0.32)	-0.76 (-2.83,1.31)	0.15 (-0.59,0.89)	EpiA						
0.43 (-0.94,1.81)	-0.53 (-1.41,0.35)	-0.10 (-2.07,1.87)	-0.06 (-2.01,1.88)	0.30 (-1.03,1.63)	1.21 (-0.30,2.71)	1.06 (-0.62,2.73)	MPD					
0.37 (-0.89,1.63)	-0.60 (-1.42,0.23)	-0.16 (-2.11,1.78)	-0.13 (-2.05,1.79)	0.23 (-1.09,1.56)	1.14 (-0.33,2.61)	0.99 (-0.65,2.64)	-0.07 (-0.74,0.61)	CD				
-0.65 (-2.10,0.80)	**-1.62 (-2.69,-0.54)**	-1.18 (-3.03,0.66)	-1.15 (-2.97,0.68)	-0.78 (-2.43,0.86)	0.12 (-1.22,1.47)	-0.03 (-1.56,1.51)	**-1.08 (-2.14,-0.02)**	-1.02 (-2.12,0.09)	EpiS			
0.17 (-1.33,1.67)	-0.80 (-1.90,0.31)	-0.36 (-2.34,1.61)	-0.33 (-2.29,1.63)	0.04 (-1.54,1.61)	0.94 (-0.57,2.46)	0.79 (-0.89,2.48)	-0.26 (-1.16,0.63)	-0.20 (-1.24,0.84)	0.82 (-0.10,1.74)	Endo		
0.68 (-1.28,2.64)	-0.28 (-1.99,1.43)	0.15 (-2.31,2.61)	0.19 (-2.25,2.63)	0.55 (-1.45,2.55)	1.46 (-0.64,3.56)	1.31 (-0.92,3.54)	0.25 (-1.40,1.90)	0.32 (-1.19,1.82)	1.33 (-0.53,3.20)	0.51 (-1.31,2.34)	CD+Fu	
0.90 (-1.53,3.33)	-0.07 (-2.30,2.17)	0.37 (-2.48,3.21)	0.40 (-2.43,3.23)	0.76 (-1.70,3.23)	1.67 (-0.87,4.22)	1.52 (-1.13,4.17)	0.47 (-1.72,2.65)	0.53 (-1.55,2.61)	1.55 (-0.80,3.90)	0.73 (-1.59,3.05)	0.21 (-1.22,1.65)	CD+ID

5. Include study with lateral LSS (Koh 2013)

Cons												
0.89 (-0.14,1.91)	ID											
0.08 (-1.44,1.59)	-0.81 (-2.43,0.81)	EpiASH										
0.25 (-1.38,1.88)	-0.63 (-2.37,1.10)	0.17 (-0.92,1.26)	BT									
0.06 (-1.37,1.48)	-0.83 (-2.23,0.56)	-0.02 (-1.97,1.92)	-0.20 (-2.24,1.84)	MSD								
-0.81 (-2.02,0.40)	**-1.70 (-3.04,-0.35)**	-0.89 (-1.79,0.01)	-1.06 (-2.15,0.03)	-0.87 (-2.59,0.86)	EpiAS							
-0.66 (-2.06,0.73)	**-1.55 (-3.06,-0.04)**	-0.74 (-1.87,0.39)	-0.92 (-2.20,0.37)	-0.72 (-2.57,1.14)	0.15 (-0.54,0.83)	EpiA						

0.36 (-0.58,1.30)	-0.53 (-1.35,0.28)	0.28 (-1.32,1.87)	0.10 (-1.60,1.81)	0.30 (-0.93,1.53)	1.16 (-0.15,2.48)	1.02 (-0.46,2.50)	MPD					
0.29 (-0.45,1.03)	-0.60 (-1.36,0.17)	0.21 (-1.34,1.76)	0.04 (-1.63,1.70)	0.24 (-1.00,1.47)	1.10 (-0.16,2.36)	0.95 (-0.48,2.39)	-0.06 (-0.69,0.56)	CD				
-0.72 (-1.87,0.43)	**-1.61 (-2.59,-0.62)**	-0.80 (-2.31,0.71)	-0.97 (-2.60,0.65)	-0.78 (-2.29,0.74)	0.09 (-1.12,1.30)	-0.06 (-1.45,1.33)	**-1.08 (-2.05,-0.10)**	**-1.01 (-2.02,-0.00)**	EpiS			
0.10 (-1.05,1.25)	-0.79 (-1.81,0.23)	0.02 (-1.60,1.64)	-0.16 (-1.89,1.58)	0.04 (-1.42,1.50)	0.91 (-0.44,2.26)	0.76 (-0.75,2.27)	-0.26 (-1.09,0.57)	-0.19 (-1.15,0.77)	0.82 (-0.03,1.66)	Endo		
0.61 (-0.97,2.19)	-0.28 (-1.87,1.31)	0.53 (-1.56,2.62)	0.35 (-1.82,2.53)	0.55 (-1.31,2.41)	1.42 (-0.46,3.30)	1.27 (-0.73,3.27)	0.25 (-1.28,1.79)	0.32 (-1.08,1.71)	1.33 (-0.40,3.05)	0.51 (-1.19,2.21)	CD+Fu	
0.82 (-1.24,2.88)	-0.07 (-2.14,2.01)	0.74 (-1.73,3.22)	0.57 (-1.98,3.11)	0.77 (-1.52,3.05)	1.63 (-0.67,3.93)	1.48 (-0.92,3.89)	0.47 (-1.56,2.49)	0.53 (-1.39,2.46)	1.54 (-0.63,3.72)	0.72 (-1.43,2.88)	0.21 (-1.11,1.54)	CD+ID

6. Exclude studies which did not mention the failure of conservative treatments in the inclusion criteria (Musacchio 2016, Kim 2013, Manchikanti 2009, Manchikanti 2012, Manchikanti 2015, Nam 2011, Karm 2018, Delitto 2015, Friedly 2016, Ghogawala 2016, Gurelik 2012, Koc 2009, Slatis 2011, Moojen 2013, Song 2016, Stromqvist 2013)

Cons						
1.93 (0.55,3.32)	ID					
0.13 (-1.18,1.44)	**-1.80 (-3.04,-0.57)**	MSD				
0.26 (-0.85,1.37)	**-1.67 (-2.50,-0.85)**	0.13 (-0.79,1.05)	MPD			
0.04 (-0.89,0.97)	**-1.89 (-2.92,-0.86)**	-0.09 (-1.01,0.83)	-0.22 (-0.83,0.39)	CD		
-0.42 (-1.79,0.96)	**-2.35 (-3.15,-1.54)**	-0.55 (-1.77,0.68)	-0.68 (-1.48,0.13)	-0.46 (-1.47,0.55)	EpiS	
0.20 (-1.07,1.47)	**-1.73 (-2.59,-0.87)**	0.07 (-1.03,1.18)	-0.05 (-0.67,0.57)	0.16 (-0.71,1.03)	0.62 (-0.03,1.27)	Endo

7. Exclude studies which did not mention typical symptoms in the inclusion criteria (Friedly 2016, Kim 2013, Karm 2018, Koc 2009, Slatis 2011, Manchikanti 2009, Manchikanti 2012, Manchikanti 2015, Nam 2011, Song 2016, Zucherman 2005)

Cons								
0.34 (-0.93,1.61)	ID							
0.03 (-1.55,1.61)	-0.31 (-1.77,1.15)	MSD						
0.05	-0.29	0.02	MPD					

(-1.12,1.21)	(-1.21,0.63)	(-1.24,1.27)						
0.05 (-0.91,1.01)	-0.29 (-1.12,0.55)	0.02 (-1.23,1.28)	0.01 (-0.66,0.67)	CD				
-0.49 (-2.31,1.32)	-0.83 (-2.50,0.83)	-0.52 (-2.40,1.35)	-0.54 (-1.93,0.85)	-0.55 (-2.09,0.99)	EpiS			
0.05 (-1.47,1.56)	-0.29 (-1.63,1.04)	0.02 (-1.57,1.60)	0.00 (-0.97,0.97)	-0.01 (-1.18,1.17)	0.54 (-0.46,1.54)	Endo		
0.37 (-1.35,2.09)	0.03 (-1.62,1.68)	0.34 (-1.56,2.24)	0.32 (-1.25,1.89)	0.32 (-1.11,1.74)	0.86 (-1.23,2.96)	0.32 (-1.52,2.17)	CD+Fu	
0.58 (-1.60,2.77)	0.24 (-1.89,2.38)	0.55 (-1.78,2.88)	0.54 (-1.53,2.61)	0.53 (-1.43,2.49)	1.08 (-1.42,3.57)	0.54 (-1.75,2.82)	0.21 (-1.14,1.56)	CD+ID

8. Exclude studies with conservative care group (Koc 2009, Slatis 2011, Weinstein 2008, Delitto 2015)

CD											
0.48 (-0.36,1.33)	ID										
-0.50 (-2.90,1.90)	-0.98 (-3.34,1.37)	EpiASH									
-0.53 (-2.91,1.85)	-1.02 (-3.35,1.32)	-0.04 (-1.29,1.22)	BT								
-0.07 (-1.41,1.27)	-0.55 (-2.08,0.97)	0.43 (-2.26,3.11)	0.46 (-2.21,3.13)	MSD							
-1.81 (-3.83,0.22)	**-2.29 (-4.26,-0.31)**	**-1.31 (-2.59,-0.02)**	**-1.27 (-2.52,-0.02)**	-1.73 (-4.09,0.62)	EpiAS						
-1.66 (-3.81,0.50)	**-2.14 (-4.25,-0.03)**	-1.16 (-2.64,0.33)	-1.12 (-2.58,0.33)	-1.58 (-4.06,0.89)	0.15 (-0.60,0.89)	EpiA					
-0.10 (-0.79,0.59)	-0.59 (-1.48,0.31)	0.40 (-1.96,2.75)	0.43 (-1.91,2.77)	-0.03 (-1.37,1.31)	1.70 (-0.27,3.68)	1.55 (-0.56,3.66)	MPD				
-1.32 (-2.54,-0.09)	**-1.80 (-2.94,-0.66)**	-0.82 (-2.88,1.24)	-0.78 (-2.82,1.26)	-1.24 (-2.96,0.48)	0.49 (-1.12,2.10)	0.34 (-1.43,2.12)	**-1.21 (-2.35,-0.07)**	EpiS			
-0.43 (-1.53,0.67)	-0.91 (-2.04,0.21)	0.07 (-2.20,2.34)	0.10 (-2.15,2.35)	-0.36 (-1.97,1.25)	1.38 (-0.49,3.24)	1.23 (-0.79,3.24)	-0.33 (-1.25,0.60)	0.89 (-0.06,1.83)	Endo		
0.32 (-1.20,1.83)	-0.17 (-1.90,1.57)	0.81 (-2.02,3.65)	0.85 (-1.97,3.67)	0.39 (-1.64,2.41)	2.12 (-0.41,4.65)	1.97 (-0.66,4.61)	0.42 (-1.25,2.09)	1.63 (-0.32,3.58)	0.75 (-1.13,2.62)	CD+Fu	
0.53 (-1.57,2.63)	0.05 (-2.22,2.31)	1.03 (-2.16,4.22)	1.06 (-2.11,4.24)	0.60 (-1.89,3.09)	2.34 (-0.58,5.25)	2.19 (-0.82,5.20)	0.63 (-1.58,2.84)	1.85 (-0.58,4.28)	0.96 (-1.41,3.33)	0.21 (-1.24,1.66)	CD+ID

9. Include studies without spinal instability (Celik 2010, Ghogawala 2016, Gurelik 2012, Haddadi 2016, Lonne 2015, Weinstein 2008, Yagi 2009)

Cons					

1.90 (-0.14,3.94)	ID				
0.18 (-1.63,1.98)	-1.72 (-3.58,0.14)	MSD			
0.35 (-1.14,1.85)	**-1.55** **(-2.94,-0.16)**	0.18 (-1.06,1.41)	MPD		
0.04 (-1.28,1.36)	**-1.86** **(-3.42,-0.30)**	-0.14 (-1.37,1.10)	-0.31 (-1.02,0.39)	CD	
0.36 (-1.56,2.27)	-1.54 (-3.63,0.55)	0.18 (-1.68,2.04)	0.00 (-1.56,1.56)	0.32 (-1.08,1.71)	CD+Fu

Long-term physical function (the numbers are presented as standardised mean difference and 95% confidence interval)

1.Exclude studies received commercial funding (Musacchio 2016, Benyamin 2016, Moojen 2015, Schmidt 2018 and Zucherman 2005)

Cons								
0.56 (-0.36,1.48)	MSD							
0.35 (-0.51,1.21)	-0.21 (-0.86,0.45)	MPD						
0.29 (-0.42,1.00)	-0.27 (-0.85,0.32)	-0.06 (-0.55,0.43)	CD					
0.32 (-0.90,1.54)	-0.24 (-1.33,0.85)	-0.03 (-0.90,0.83)	0.03 (-0.97,1.02)	Endo				
-0.52 (-2.00,0.96)	-1.08 (-2.50,0.34)	-0.87 (-2.26,0.51)	-0.81 (-2.11,0.48)	-0.84 (-2.48,0.80)	CD+ID			
0.40 (-0.59,1.39)	-0.16 (-1.03,0.72)	0.05 (-0.69,0.79)	0.11 (-0.58,0.80)	0.08 (-1.06,1.22)	0.92 (-0.55,2.39)	CD+Fu		
-0.53 (-2.20,1.15)	-1.09 (-2.69,0.52)	-0.88 (-2.41,0.65)	-0.82 (-2.34,0.69)	-0.85 (-2.60,0.91)	-0.01 (-2.00,1.99)	-0.93 (-2.48,0.62)	EpiS	
0.81 (-0.26,1.88)	0.25 (-0.71,1.20)	0.45 (-0.36,1.27)	0.51 (-0.29,1.31)	0.49 (-0.70,1.68)	1.33 (-0.20,2.85)	0.40 (-0.46,1.27)	**1.33** **(0.05,2.62)**	ID

2.Exclude studies with level 1 only (Yagi 2009)

Cons								
0.48 (-0.31,1.26)	MSD							

0.09 (-0.65,0.83)	-0.38 (-0.96,0.19)	MPD						
0.29 (-0.31,0.89)	-0.18 (-0.69,0.32)	0.20 (-0.23,0.63)	CD					
-0.06 (-1.02,0.90)	-0.54 (-1.38,0.31)	-0.15 (-0.81,0.50)	-0.35 (-1.10,0.40)	Endo				
0.16 (-0.71,1.04)	-0.31 (-1.12,0.50)	0.07 (-0.67,0.82)	-0.13 (-0.76,0.51)	0.23 (-0.74,1.19)	CD+ID			
0.23 (-0.57,1.03)	-0.25 (-0.96,0.46)	0.14 (-0.47,0.74)	-0.06 (-0.59,0.47)	0.29 (-0.56,1.14)	0.06 (-0.63,0.76)	CD+Fu		
-1.00 **(-2.00,-0.01)**	**-1.48** **(-2.39,-0.57)**	**-1.10** **(-1.88,-0.31)**	**-1.30** **(-2.09,-0.50)**	**-0.94** **(-1.73,-0.15)**	**-1.17** **(-2.16,-0.17)**	**-1.23** **(-2.11,-0.35)**	EpiS	
0.61 (-0.20,1.43)	0.14 (-0.58,0.85)	0.52 (-0.06,1.11)	0.32 (-0.23,0.87)	0.67 (-0.10,1.45)	0.45 (-0.36,1.26)	0.38 (-0.27,1.04)	**1.62** **(0.94,2.29)**	ID

3.Exclude studies with mixed type of disease (Azzazi 2010, Benyamin 2016, Mobbs 2014, Weinstein 2008)

Cons								
0.63 (-0.46,1.72)	MSD							
0.35 (-0.70,1.40)	-0.28 (-0.97,0.40)	MPD						
0.40 (-0.51,1.30)	-0.23 (-0.84,0.37)	0.05 (-0.48,0.58)	CD					
0.32 (-1.07,1.70)	-0.31 (-1.44,0.81)	-0.03 (-0.93,0.86)	-0.08 (-1.13,0.96)	Endo				
0.25 (-0.94,1.45)	-0.38 (-1.35,0.60)	-0.10 (-1.01,0.82)	-0.14 (-0.92,0.63)	-0.06 (-1.34,1.22)	CD+ID			
0.32 (-0.82,1.46)	-0.31 (-1.21,0.58)	-0.03 (-0.82,0.76)	-0.08 (-0.77,0.61)	0.00 (-1.19,1.20)	0.06 (-0.80,0.93)	CD+Fu		
-0.98 (-2.48,0.53)	**-1.61** **(-2.93,-0.29)**	**-1.32** **(-2.57,-0.08)**	**-1.37** **(-2.57,-0.18)**	-1.29 (-2.82,0.24)	-1.23 (-2.65,0.19)	-1.29 (-2.66,0.08)	EpiS	
0.76 (-0.42,1.95)	0.13 (-0.82,1.08)	0.41 (-0.42,1.25)	0.36 (-0.40,1.13)	0.45 (-0.78,1.67)	0.51 (-0.58,1.59)	0.44 (-0.57,1.46)	**1.74** **(0.82,2.66)**	ID

4.Studies with more than 20% lost to follow up (Mobbs 2014)

Cons								
0.51 (-0.37,1.39)	MSD							

0.21 (-0.63,1.04)	-0.31 (-0.94,0.33)	MPD						
0.29 (-0.38,0.97)	-0.22 (-0.78,0.34)	0.09 (-0.40,0.57)	CD					
0.04 (-1.04,1.12)	-0.47 (-1.42,0.47)	-0.17 (-0.90,0.57)	-0.25 (-1.09,0.58)	Endo				
0.17 (-0.81,1.15)	-0.34 (-1.24,0.56)	-0.03 (-0.87,0.80)	-0.12 (-0.84,0.59)	0.13 (-0.95,1.21)	CD+ID			
0.26 (-0.64,1.16)	-0.25 (-1.04,0.54)	0.06 (-0.62,0.73)	-0.03 (-0.62,0.56)	0.22 (-0.73,1.18)	0.09 (-0.69,0.87)	CD+Fu		
-0.94 (-2.05,0.18)	**-1.45 (-2.46,-0.44)**	**-1.14 (-2.02,-0.27)**	**-1.23 (-2.12,-0.34)**	**-0.97 (-1.86,-0.09)**	-1.11 (-2.22,0.00)	**-1.20 (-2.18,-0.22)**	EpiS	
0.66 (-0.25,1.57)	0.15 (-0.65,0.94)	0.46 (-0.20,1.11)	0.37 (-0.24,0.98)	0.62 (-0.24,1.49)	0.49 (-0.41,1.39)	0.40 (-0.33,1.13)	**1.60 (0.85,2.35)**	ID

5. Exclude studies which has drug therapy in the conservative care group (Slatis 2011 and Weinstein 2008)

Cons								
0.39 (-0.96,1.74)	MSD							
0.14 (-1.17,1.44)	-0.26 (-0.90,0.39)	MPD						
0.15 (-1.07,1.36)	-0.24 (-0.83,0.34)	0.01 (-0.46,0.48)	CD					
-0.04 (-1.53,1.45)	-0.43 (-1.41,0.55)	-0.18 (-0.94,0.59)	-0.19 (-1.05,0.67)	Endo				
0.03 (-1.39,1.46)	-0.36 (-1.30,0.57)	-0.11 (-0.96,0.75)	-0.12 (-0.86,0.63)	0.07 (-1.04,1.19)	CD+ID			
0.14 (-1.22,1.50)	-0.25 (-1.07,0.57)	0.00 (-0.69,0.69)	-0.01 (-0.62,0.60)	0.18 (-0.81,1.17)	0.11 (-0.71,0.92)	CD+Fu		
-1.04 (-2.56,0.49)	**-1.43 (-2.48,-0.38)**	**-1.17 (-2.08,-0.27)**	**-1.19 (-2.11,-0.27)**	**-1.00 (-1.92,-0.07)**	-1.07 (-2.22,0.09)	**-1.18 (-2.20,-0.16)**	EpiS	
0.55 (-0.82,1.92)	0.16 (-0.67,0.98)	0.41 (-0.26,1.08)	0.40 (-0.23,1.03)	0.59 (-0.31,1.49)	0.52 (-0.42,1.46)	0.41 (-0.35,1.17)	**1.59 (0.80,2.37)**	ID

6. Exclude studies which did not mention the failure of conservative treatments in the inclusion criteria (Forsth 2016, Ghogawala 2016, Liu 2013, Slatis 2011, Mobbs 2014, Moojen 2013, Stromqvist 2013, Zucherman 2005)

Cons								
0.27	MSD							

(-1.04,1.59)								
0.34 (-0.96,1.63)	0.06 (-0.71,0.84)	MPD						
0.10 (-1.07,1.26)	-0.18 (-0.79,0.44)	-0.24 (-0.81,0.34)	CD					
0.32 (-1.18,1.81)	0.04 (-1.03,1.12)	-0.02 (-0.77,0.74)	0.22 (-0.72,1.16)	Endo				
0.13 (-1.26,1.52)	-0.14 (-1.11,0.82)	-0.20 (-1.07,0.66)	0.03 (-0.73,0.79)	-0.19 (-1.31,0.94)	CD+ID			
0.51 (-0.97,1.99)	0.24 (-0.83,1.30)	0.18 (-0.67,1.02)	0.41 (-0.50,1.33)	0.20 (-0.88,1.27)	0.38 (-0.53,1.29)	CD+Fu		
-0.37 (-1.97,1.23)	-0.64 (-1.86,0.58)	-0.70 (-1.67,0.26)	-0.47 (-1.57,0.63)	-0.69 (-1.62,0.25)	-0.50 (-1.73,0.73)	-0.88 (-2.00,0.23)	EpiS	
1.39 (-0.14,2.91)	1.11 (-0.01,2.23)	**1.05 (0.20,1.90)**	**1.29 (0.30,2.27)**	**1.07 (0.10,2.04)**	**1.26 (0.15,2.36)**	0.87 (-0.06,1.81)	**1.76 (0.98,2.53)**	ID

7. Exclude studies which did not mention typical symptoms in the inclusion criteria (Azzazi 2010, Cho 2007, Forsth 2016, Liu 2013, Slatis 2011, Marsh 2014, Schmidt 2018, Zucherman 2005)

Cons								
-0.01 (-1.12,1.09)	MSD							
0.19 (-0.79,1.17)	0.21 (-0.60,1.01)	MPD						
0.12 (-0.72,0.97)	0.14 (-0.57,0.85)	-0.07 (-0.56,0.43)	CD					
0.11 (-1.11,1.34)	0.13 (-0.97,1.23)	-0.08 (-0.85,0.70)	-0.01 (-0.90,0.88)	Endo				
0.41 (-1.32,2.14)	0.42 (-1.23,2.08)	0.22 (-1.30,1.73)	0.29 (-1.23,1.80)	0.29 (-1.40,1.99)	CD+ID			
0.25 (-1.01,1.52)	0.27 (-0.89,1.43)	0.06 (-0.88,1.01)	0.13 (-0.81,1.07)	0.14 (-1.07,1.35)	-0.15 (-1.34,1.03)	CD+Fu		
-0.69 (-2.06,0.68)	-0.67 (-1.94,0.59)	-0.88 (-1.92,0.17)	-0.81 (-1.89,0.27)	-0.80 (-1.78,0.18)	-1.10 (-2.91,0.72)	-0.94 (-2.32,0.44)	EpiS	
0.55 (-0.54,1.64)	0.57 (-0.41,1.54)	0.36 (-0.37,1.09)	0.43 (-0.27,1.13)	0.44 (-0.53,1.40)	0.14 (-1.49,1.78)	0.30 (-0.83,1.43)	**1.24 (0.23,2.24)**	ID

8. Exclude studies with conservative care group (Slatis 2011, Westein 2008, Delitto 2015)

CD							
0.24 (-0.34,0.83)	MSD						

-0.01 (-0.48,0.46)	-0.26 (-0.90,0.39)	MPD					
-0.19 (-1.05,0.67)	-0.43 (-1.41,0.55)	-0.18 (-0.94,0.59)	Endo				
-0.12 (-0.86,0.63)	-0.36 (-1.30,0.57)	-0.11 (-0.96,0.75)	0.07 (-1.04,1.19)	CD+ID			
-0.01 (-0.62,0.60)	-0.25 (-1.07,0.57)	0.00 (-0.69,0.69)	0.18 (-0.81,1.17)	0.11 (-0.71,0.92)	CD+Fu		
-1.19 **(-2.11,-0.27)**	**-1.43** **(-2.48,-0.38)**	**-1.17** **(-2.08,-0.27)**	**-1.00** **(-1.92,-0.07)**	-1.07 (-2.22,0.09)	**-1.18** **(-2.20,-0.16)**	EpiS	
0.40 (-0.23,1.03)	0.16 (-0.67,0.98)	0.41 (-0.26,1.08)	0.59 (-0.31,1.49)	0.52 (-0.42,1.46)	0.41 (-0.35,1.17)	**1.59** **(0.80,2.37)**	ID

9. Include studies without spinal instability (Celik 2010, Cho 2007, Forsth 2016, Ghogawala 2016, Haddadi 2016, Liu 2013, Lonne 2015, Marsh 2014, Mobbs 2014, Rajasekaran 2013, Schmidt 2018, Thom 2005, Weinstein 2008, Yagi 2009)

Cons						
0.58 (-0.66,1.82)	MSD					
0.42 (-0.77,1.62)	-0.15 (-0.80,0.49)	MPD				
0.10 (-0.99,1.19)	-0.48 (-1.08,0.12)	-0.33 (-0.82,0.17)	CD			
-0.15 (-1.51,1.20)	-0.73 (-1.74,0.27)	-0.58 (-1.53,0.37)	-0.25 (-1.06,0.55)	CD+ID		
0.08 (-1.27,1.43)	-0.50 (-1.50,0.50)	-0.34 (-1.29,0.60)	-0.02 (-0.82,0.79)	0.23 (-0.90,1.37)	CD+Fu	
2.22 **(0.54,3.90)**	**1.64** **(0.29,2.99)**	**1.79** **(0.61,2.98)**	**2.12** **(0.84,3.40)**	**2.37** **(0.85,3.89)**	**2.14** **(0.62,3.65)**	ID

All-cause mortality (the numbers are presented as odds ratio and 95% confidence interval)

1.Exclude studies received commercial funding (Ghogawala 2016)

Cons					
0.25 (0.01,6.98)	CD+ID				
0.26 (0.01,4.58)	1.02 (0.01,70.13)	MSD			

0.26 (0.02,2.96)	1.02 (0.02,52.78)	1.00 (0.11,9.30)	MPD		
2.81 (0.80,9.90)	11.16 (0.33,376.39)	10.90 (0.49,242.13)	10.92 (0.73,162.89)	CD+Fu	
0.80 (0.34,1.86)	3.16 (0.13,78.51)	3.09 (0.20,48.23)	3.09 (0.31,30.54)	0.28 (0.07,1.19)	CD

2. Exclude studies with mixed type of disease (Amundsen 2000, Weinstein 2007, Weinstein 2008)

Cons					
0.25 (0.01,7.62)	CD+ID				
0.25 (0.01,5.08)	1.02 (0.01,70.13)	MSD			
0.25 (0.02,3.35)	1.02 (0.02,52.78)	1.00 (0.11,9.30)	MPD		
4.41 (0.38,51.16)	17.79 (0.38,844.18)	17.37 (0.53,565.88)	17.40 (0.76,399.79)	CD+Fu	
0.78 (0.24,2.59)	3.16 (0.13,78.51)	3.09 (0.20,48.23)	3.09 (0.31,30.54)	0.18 (0.02,1.51)	CD

3. Exclude studies which has drug therapy in the conservative care group (Slatis 2011, Weinstein 2007, Weinstein 2008)

Cons					
0.14 (0.00,4.73)	CD+ID				
0.14 (0.01,3.19)	1.02 (0.01,70.13)	MSD			
0.14 (0.01,2.14)	1.02 (0.02,52.78)	1.00 (0.11,9.30)	MPD		
2.51 (0.19,32.80)	17.79 (0.38,844.18)	17.37 (0.53,565.88)	17.40 (0.76,399.79)	CD+Fu	
0.45 (0.11,1.85)	3.16 (0.13,78.51)	3.09 (0.20,48.23)	3.09 (0.31,30.54)	0.18 (0.02,1.51)	CD

4. Exclude studies which did not mention the failure of conservative treatments in the inclusion criteria (Amundsen 2000, Delitto 2015, Forsth 2016, Ghogawala 2016, Ko 2019, Slatis 2011)

Cons					
0.39 (0.00,3.88e+08)	CD+ID				
0.40 (0.00,6.28e+16)	1.03 (0.00,3.20e+08)	MSD			
0.40 (0.00,6.28e+16)	1.03 (0.00,3.20e+08)	1.00 (0.10,10.03)	MPD		
2.64 (0.02,446.36)	6.82 (0.00,5.69e+07)	6.63 (0.00,7.83e+15)	6.63 (0.00,7.83e+15)	CD+Fu	
1.22 (0.00,5.43e+06)	3.16 (0.01,1415.72)	3.08 (0.00,1.25e+11)	3.08 (0.00,1.25e+11)	0.46 (0.00,16626.80)	CD

5. Exclude studies which did not mention typical symptoms in the inclusion criteria (Amundsen 2000, Forsth 2016, Schmidt 2018, Slatis 2011)

Cons				
0.25 (0.01,4.82)	MSD			
0.25 (0.02,3.16)	1.00 (0.11,9.30)	MPD		
2.95 (0.84,10.43)	11.71 (0.50,272.35)	11.73 (0.75,184.44)	CD+Fu	
0.78 (0.27,2.29)	3.09 (0.20,48.23)	3.09 (0.31,30.54)	0.26 (0.06,1.22)	CD

6. Exclude studies with conservative care group (Amundsen 2000, Delitto 2015, Slatis 2011, Weinstein 2008, Weinstein 2007)

CD				

0.32 (0.01,7.84)	CD+ID			
0.32 (0.02,5.05)	1.02 (0.01,70.13)	MSD		
0.32 (0.03,3.19)	1.02 (0.02,52.78)	1.00 (0.11,9.30)	MPD	
5.62 (0.66,47.84)	17.79 (0.38,844.18)	17.37 (0.53,565.88)	17.40 (0.76,399.79)	CD+Fu

7. Include studies without spinal instability (Forsth 2016, Ghogawala 2016, Haddadi 2016, Ko 2019, Schmidt 2018, Weinstein 2008)

Cons					
0.39 (0.01,13.48)	CD+ID				
0.40 (0.02,9.14)	1.02 (0.01,70.13)	MSD			
0.40 (0.03,6.16)	1.02 (0.02,52.78)	1.00 (0.11,9.30)	MPD		
6.89 (0.50,94.84)	17.79 (0.38,844.18)	17.37 (0.53,565.88)	17.40 (0.76,399.79)	CD+Fu	
1.22 (0.27,5.57)	3.16 (0.13,78.51)	3.09 (0.20,48.23)	3.09 (0.31,30.54)	0.18 (0.02,1.51)	CD

8. Exclude the old study (Amundsen 2000)

Cons					
0.27 (0.01,7.56)	CD+ID				
0.28 (0.02,4.97)	1.02 (0.01,70.13)	MSD			

0.28 (0.02,3.22)	1.02 (0.02,52.78)	1.00 (0.11,9.30)	MPD		
3.14 (0.96,10.26)	11.61 (0.36,372.74)	11.33 (0.54,238.19)	11.35 (0.81,158.90)	CD+Fu	
0.86 (0.36,2.07)	3.16 (0.13,78.51)	3.09 (0.20,48.23)	3.09 (0.31,30.54)	0.27 (0.07,1.01)	CD

Appendix S14: Additional information about secondary outcomes

Back pain (0-10-point Visual Analog Scale)

For short-term back pain, nine randomised controlled trials including 1003 patients and seven interventions (six surgical interventions and conservative care) were included. For long-term back pain, twelve randomised controlled trials including 1403 patients and seven interventions (six surgical interventions and conservative care) were included.

Leg Pain (0-10-point Visual Analog Scale)

Nine randomised controlled trials, with 976 patients and six interventions (five surgical interventions and conservative care) provided data on short term follow up assessments. For long-term leg pain, 15 randomised controlled trials including 1614 patients and eight interventions (seven surgical interventions and conservative care) were included.

Overall pain (0-10-point Visual Analog Scale)

For short-term overall pain, nine randomised controlled trials including 806 patients and seven interventions (five epidural injection interventions, one surgical intervention and conservative care) were included.

Global impression of recovery

For short-term global impression of recovery, six randomised controlled trials including 965 patients and six interventions (four surgical interventions, one epidural injection intervention and conservative care) were included. For long-term global impression of recovery, seven randomised controlled trials including 958 patients and six interventions (five surgical interventions and conservative care) were included.

Mobility

Three randomised controlled trials including 217 patients and three interventions (two surgical interventions and conservative care) were included in the short-term mobility analyses and five trials with 558 patients and four interventions (three surgical interventions and conservative care) in the long-term mobility analyses.

Treatment withdrawal (due to any reason)

Eighteen randomised controlled trials including 2831 patients and nine interventions (seven surgical interventions, one epidural injection intervention and conservative care) were included.

Adverse effect

For adverse effect due to any reason, 26 randomised controlled trials including 2811 patients and nine interventions (seven surgical interventions, one epidural injection intervention and conservative care) were included. For intervention related adverse effect, 26 randomised controlled trials including 3092 patients and nine interventions (seven surgical interventions, one epidural injection intervention and conservative care) were included.

Reoperation rate

Twelve randomised controlled trials including 1391 patients and six interventions (six surgical interventions) were included.

Appendix S15: Ranking

For two primary outcomes for effectiveness (short-term physical function and long-term physical function), the SUCRA results showed that interspinous device, midline preserving decompression, conventional open decompression and conventional open decompression with fusion were likely to be the most effective, with the average probabilities of more than 50%. For the primary outcome of safety (all-cause mortality), only six interventions were included: conservative care, conventional open decompression with fusion and conventional open decompression were likely to be the safest. Overall, however, treatment effects were too small to be considered significant.

Short-term physical function

Intervention	SUCRA	Mean rank
ID	84.1	2.9
CD+ID	72	4.4
CD+Fu	67.1	4.9
EpiASH	64.7	5.2
BT	63.4	5.4
MPD	61.1	5.7
CD	57.3	6.1
Endo	47.5	7.3
MSD	46.1	7.5
Cons	41.9	8.0
EpiA	19.0	10.7
EpiS	14	11.3
EpiAS	11.7	11.6

Long-term physical function

Intervention	SUCRA	Mean rank
ID	87.9	2.0
MSD	77.5	2.8
CD	55.5	4.6
CD+Fu	54.3	4.7
MPD	53.8	4.7
CD+ID	45.9	5.3
Endo	40.6	5.7
Cons	33.0	6.4

EpiS	1.4	8.9

Short-term back pain

Intervention	SUCRA	Mean rank
MPD	71.1	2.7
ID	65.8	3.1
MSD	65.1	3.1
CD	56.2	3.6
CD+ID	42.7	4.4
CD+Fu	33.7	5.0
Cons	15.4	6.1

Long-term back pain

Intervention	SUCRA	Mean rank
MSD	72.5	2.6
ID	67.1	3.0
MPD	63.6	3.2
CD	51.6	3.9
CD+ID	46.7	4.2
CD+Fu	40.1	4.6
Cons	8.4	6.5

Short-term leg pain

Intervention	SUCRA	Mean rank
MSD	83.5	1.8
ID	64	2.8
Endo	58.2	3.1
CD	48.2	3.6
MPD	42.2	3.9
Cons	3.8	5.8

Long-term leg pain

Intervention	SUCRA	Mean rank
MSD	70.6	3.1
ID	65.7	3.4
CD	53.9	4.2

MPD	53.1	4.3
Endo	53.0	4.3
CD+ID	48.5	4.6
CD+Fu	46.6	4.7
Cons	8.6	7.4

Short-term overall pain

Intervention	SUCRA	Mean rank
EpiASC	96.7	1.2
EpiASH	82.3	2.1
Endo	65.3	3.1
Cons	35.3	4.9
EpiAS	34.1	5.0
EpiA	27.2	5.4
EpiS	9.3	6.4

Short-term global impression of recovery

Intervention	SUCRA	Mean rank
Endo	96.2	1.2
MPD	75.7	2.2
EpiS	44.6	3.8
CD	39.0	4.0
MSD	36.7	4.2
Cons	7.8	5.6

Long-term global impression of recovery

Intervention	SUCRA	Mean rank
MPD	89.2	1.5
MSD	73.5	2.3
ID	48.3	3.6
CD	37.8	4.1
CD+Fu	36.3	4.2
Cons	14.9	5.3

Short-term mobility

Intervention	SUCRA	Mean rank

MPD	85.3	1.3
Cons	35.3	2.3
CD	29.4	2.4

Long-term mobility

Intervention	SUCRA	Mean rank
MPD	86.3	1.4
CD	63.6	2.1
CD+Fu	26.9	3.2
Cons	23.2	3.3

All-cause mortality

Intervention	SUCRA	Mean rank
CD+Fu	95.8	1.2
Cons	64.9	2.8
CD	53.2	3.3
CD+ID	30.3	4.5
MSD	28.8	4.6
MPD	27.0	4.6

Adverse effect due to any reason

Intervention	SUCRA	Mean rank
Endo	91.0	1.7
EpiS	71.8	3.3
MPD	71.3	3.3
Cons	64.3	3.9
ID	63.8	3.9
MSD	28.6	6.7
CD	24.2	7.1
CD+ID	22.5	7.2
CD+Fu	12.5	8.0

Intervention related adverse effect

Intervention	SUCRA	Mean rank
Cons	97.1	1.2
Endo	79.3	2.7

ID	66.5	3.7
EpiS	62.9	4.0
MPD	61.3	4.1
CD+ID	28.5	6.7
MSD	25.8	6.9
CD	20.2	7.4
CD+Fu	8.4	8.3

Reoperation rate

Intervention	SUCRA	Mean rank
MPD	86.7	1.7
Endo	69.9	2.5
CD	49.4	3.5
CD+Fu	47.2	3.6
MSD	46.2	3.7
ID	0.6	6.0

Treatment withdrawal due to any reason

Intervention	SUCRA	Mean rank
Endo	92.3	1.6
MPD	72.4	3.2
CD	51.7	4.9
ID	50.6	5.0
Cons	43.5	5.5
CD+ID	42.6	5.6
MSD	42.1	5.6
EpiS	39.8	5.8
CD+Fu	15.0	7.8

References

1. Amundsen T, Weber H, Nordal HJ, Magnaes B, Abdelnoor M, Lilleas F. Lumbar spinal stenosis: conservative or surgical management?: A prospective 10-year study. *Spine.* 2000;25(11):1424-1435; discussion 1435-1426.

2. Anderson PA, Tribus CB, Kitchel SH. Treatment of neurogenic claudication by interspinous decompression: application of the X STOP device in patients with lumbar degenerative spondylolisthesis. *Journal of neurosurgery Spine.* 2006;4(6):463-471.

3. Azzazi A, Elhawary Y. Dynamic stabilization using X-stop versus transpedicular screw fixation in the treatment of lumbar canal stenosis; comparative study of the clinical outcome. *Neurosurgery Quarterly.* 2010;20(3):165-169.

4. Benyamin RM, Staats PS, Mi DASEI. MILD(R) Is an Effective Treatment for Lumbar Spinal Stenosis with Neurogenic Claudication: MiDAS ENCORE Randomized Controlled Trial. *Pain physician.* 2016;19(4):229-242.

5. Brown LL. A double-blind, randomized, prospective study of epidural steroid injection vs. the mild(R) procedure in patients with symptomatic lumbar spinal stenosis. *Pain practice : the official journal of World Institute of Pain.* 2012;12(5):333-341.

6. Cavusoglu H, Kaya RA, Turkmenoglu ON, Tuncer C, Colak I, Aydin Y. Midterm outcome after unilateral approach for bilateral decompression of lumbar spinal stenosis: 5-year prospective study. *European spine journal : official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society.* 2007;16(12):2133-2142.

7. Celik SE, Celik S, Goksu K, Kara A, Ince I. Microdecompressive laminatomy with a 5-year follow-up period for severe lumbar spinal stenosis. *Journal of spinal disorders & techniques.* 2010;23(4):229-235.

8. Cho DY, Lin HL, Lee WY, Lee HC. Split-spinous process laminotomy and discectomy for degenerative lumbar spinal stenosis: a preliminary report. *Journal of neurosurgery Spine.* 2007;6(3):229-239.

9. Delitto A, Piva SR, Moore CG, et al. Surgery versus nonsurgical treatment of lumbar spinal stenosis: a randomized trial. *Ann Intern Med.* 2015;162(7):465-473.

10. Elsheikh NA, Amr YM. Effect of Adding Calcitonin to Translaminar Epidural Steroid in Degenerative Lumbar Spinal Canal Stenosis. *Pain physician.* 2016;19(3):139-146.

11. Forsth P, Olafsson G, Carlsson T, et al. A Randomized, Controlled Trial of Fusion Surgery for Lumbar Spinal Stenosis. *The New England journal of medicine.* 2016;374(15):1413-1423.

12. Friedly JL, Comstock BA, Turner JA, et al. Long-Term Effects of Repeated Injections of Local Anesthetic With or Without Corticosteroid for Lumbar Spinal Stenosis: A Randomized Trial. *Archives of physical medicine and rehabilitation.* 2017;98(8):1499-1507.e1492.

13. Friedly JL, Comstock BA, Turner JA, et al. A randomized trial of epidural glucocorticoid injections for spinal stenosis. *The New England journal of medicine.* 2014;371(1):11-21.

14. Ghogawala Z, Dziura J, Butler WE, et al. Laminectomy plus Fusion versus Laminectomy Alone for Lumbar Spondylolisthesis. *The New England journal of medicine.* 2016;374(15):1424-1434.

15. Grob D, Humke T, Dvorak J. Degenerative lumbar spinal stenosis. Decompression with and without arthrodesis. *The Journal of bone and joint surgery American volume.* 1995;77(7):1036-1041.

16. Gurelik M, Bozkina C, Zafer K, Karadag O, Unal O, Bayrakli F. Unilateral laminotomy for decompression of lumbar stenosis is effective and safe: a prospective randomized comparative study. *Journal of Neurological Sciences (Turkish).* 2012;29(4):744-753.

17. Haddadi K, Ganjeh Qazvini HR. Outcome after Surgery of Lumbar Spinal Stenosis: A Randomized Comparison of Bilateral Laminotomy, Trumpet Laminectomy, and Conventional Laminectomy. *Frontiers in surgery.* 2016;3:19.

18. Hallett A, Huntley JS, Gibson JN. Foraminal stenosis and single-level degenerative disc disease: a randomized controlled trial comparing decompression with decompression and instrumented fusion. *Spine.* 2007;32(13):1375-1380.

19. Kang T, Park SY, Kang CH, Lee SH, Park JH, Suh SW. Is biportal technique/endoscopic spinal surgery satisfactory for lumbar spinal stenosis patients?: A prospective randomized comparative study. *Medicine.* 2019;98(18):e15451.

20. Karm MH, Choi SS, Kim DH, et al. Percutaneous Epidural Adhesiolysis Using Inflatable Balloon Catheter and Balloon-less Catheter in Central Lumbar Spinal Stenosis with Neurogenic Claudication: A Randomized Controlled Trial. *Pain physician.* 2018;21(6):593-606.

21. Kim SH, Choi WJ, Suh JH, et al. Effects of transforaminal balloon treatment in patients with lumbar foraminal stenosis: a randomized, controlled, double-blind trial. *Pain physician.* 2013;16(3):213-224.

22. Ko S, Oh T. Comparison of bilateral decompression via unilateral laminotomy and conventional laminectomy for single-level degenerative lumbar spinal stenosis regarding low back pain, functional outcome, and quality of life - A Randomized Controlled, Prospective Trial. *Journal of orthopaedic surgery and research.* 2019;14(1):252.

23. Koc Z, Ozcakir S, Sivrioglu K, Gurbet A, Kucukoglu S. Effectiveness of physical therapy and epidural steroid injections in lumbar spinal stenosis. *Spine.* 2009;34(10):985-989.

24. Koh WU, Choi SS, Park SY, et al. Transforaminal hypertonic saline for the treatment of lumbar lateral canal stenosis: a double-blinded, randomized, active-control trial. *Pain physician.* 2013;16(3):197-211.

25. Komp M, Hahn P, Oezdemir S, et al. Bilateral spinal decompression of lumbar central stenosis with the full-endoscopic interlaminar versus microsurgical laminotomy technique: a prospective, randomized, controlled study. *Pain physician.* 2015;18(1):61-70.

26. Liu X, Yuan S, Tian Y. Modified unilateral laminotomy for bilateral decompression for lumbar spinal stenosis: technical note. *Spine.* 2013;38(12):E732-737.

27. Lonne G, Johnsen LG, Rossvoll I, et al. Minimally invasive decompression versus x-stop in lumbar spinal stenosis: a randomized controlled multicenter study. *Spine.* 2015;40(2):77-85.

28. Fukusaki M, Kobayashi I, Hara T, Sumikawa K. Symptoms of spinal stenosis do not improve after epidural steroid injection. *The Clinical journal of pain.* 1998;14(2):148-151.

29. Manchikanti L, Cash KA, McManus CD, Pampati V, Singh V, Benyamin R. The preliminary results of a comparative effectiveness evaluation of adhesiolysis and caudal epidural

injections in managing chronic low back pain secondary to spinal stenosis: a randomized, equivalence controlled trial. *Pain physician.* 2009;12(6):E341-354.

30. Manchikanti L, Cash KA, McManus CD, Pampati V, Abdi S. Preliminary results of a randomized, equivalence trial of fluoroscopic caudal epidural injections in managing chronic low back pain: Part 4--Spinal stenosis. *Pain physician.* 2008;11(6):833-848.

31. Manchikanti L, Cash KA, McManus CD, Pampati V, Fellows B. Fluoroscopic caudal epidural injections with or without steroids in managing pain of lumbar spinal stenosis: one-year results of randomized, double-blind, active-controlled trial. *Journal of spinal disorders & techniques.* 2012;25(4):226-234.

32. Manchikanti L, Cash KA, McManus CD, Pampati V, Fellows B. Results of 2-year follow-up of a randomized, double-blind, controlled trial of fluoroscopic caudal epidural injections in central spinal stenosis. *Pain physician.* 2012;15(5):371-384.

33. Manchikanti L, Cash KA, McManus CD, Damron KS, Pampati V, Falco FJ. Lumbar interlaminar epidural injections in central spinal stenosis: preliminary results of a randomized, double-blind, active control trial. *Pain physician.* 2012;15(1):51-63.

34. Manchikanti L, Cash KA, McManus CD, Damron KS, Pampati V, Falco FJ. A randomized, double-blind controlled trial of lumbar interlaminar epidural injections in central spinal stenosis: 2-year follow-up. *Pain physician.* 2015;18(1):79-92.

35. Marsh GD, Mahir S, Leyte A. A prospective randomised controlled trial to assess the efficacy of dynamic stabilisation of the lumbar spine with the Wallis ligament. *European spine journal : official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society.* 2014;23(10):2156-2160.

36. Mobbs RJ, Li J, Sivabalan P, Raley D, Rao PJ. Outcomes after decompressive laminectomy for lumbar spinal stenosis: comparison between minimally invasive unilateral laminectomy for bilateral decompression and open laminectomy: clinical article. *Journal of neurosurgery Spine.* 2014;21(2):179-186.

37. Moojen WA, Arts MP, Jacobs WC, et al. Interspinous process device versus standard conventional surgical decompression for lumbar spinal stenosis: randomized controlled trial. *Bmj*. 2013;347:f6415.

38. Moojen WA, Arts MP, Jacobs WC, et al. IPD without bony decompression versus conventional surgical decompression for lumbar spinal stenosis: 2-year results of a double-blind randomized controlled trial. *European spine journal : official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society.* 2015;24(10):2295-2305.

39. Bae HW, Davis RJ, Lauryssen C, Leary S, Maislin G, Musacchio MJ, Jr. Three-Year Follow-up of the Prospective, Randomized, Controlled Trial of Coflex Interlaminar Stabilization vs Instrumented Fusion in Patients With Lumbar Stenosis. *Neurosurgery.* 2016;79(2):169-181.

40. Davis RJ, Errico TJ, Bae H, Auerbach JD. Decompression and Coflex interlaminar stabilization compared with decompression and instrumented spinal fusion for spinal stenosis and low-grade degenerative spondylolisthesis: two-year results from the prospective, randomized, multicenter, Food and Drug Administration Investigational Device Exemption trial. *Spine.* 2013;38(18):1529-1539.

41. Musacchio MJ, Lauryssen C, Davis RJ, et al. Evaluation of Decompression and Interlaminar Stabilization Compared with Decompression and Fusion for the Treatment of Lumbar Spinal Stenosis: 5-year Follow-up of a Prospective, Randomized, Controlled Trial. *International journal of spine surgery.* 2016;10:6.

42. Nam H-S, Park YB. Effects of transforaminal injection for degenerative lumbar scoliosis combined with spinal stenosis. *Annals of rehabilitation medicine.* 2011;35(4):514.

43. Park SM, Kim GU, Kim HJ, et al. Is the Use of a Unilateral Biportal Endoscopic Approach Associated with Rapid Recovery After Lumbar Decompressive Laminectomy? A Preliminary Analysis of a Prospective Randomized Controlled Trial. *World neurosurgery.* 2019;128:e709-e718.

44. Rajasekaran S, Thomas A, Kanna RM, Prasad Shetty A. Lumbar spinous process splitting decompression provides equivalent outcomes to conventional midline decompression in

degenerative lumbar canal stenosis: a prospective, randomized controlled study of 51 patients. *Spine.* 2013;38(20):1737-1743.

45. Ruetten S, Komp M, Merk H, Godolias G. Surgical treatment for lumbar lateral recess stenosis with the full-endoscopic interlaminar approach versus conventional microsurgical technique: a prospective, randomized, controlled study. *Journal of neurosurgery Spine.* 2009;10(5):476-485.

46. Schmidt S, Franke J, Rauschmann M, Adelt D, Bonsanto MM, Sola S. Prospective, randomized, multicenter study with 2-year follow-up to compare the performance of decompression with and without interlaminar stabilization. *Journal of neurosurgery Spine.* 2018;28(4):406-415.

47. Skoro I, Stancic M, Kovacevic M, Duric KS. Long-Term Results and Efficacy of Laminectomy with Fusion Versus Young Laminoplasty for the Treatment of Degenerative Spinal Stenosis. *World neurosurgery.* 2016;89:387-392.

48. Malmivaara A, Slatis P, Heliovaara M, et al. Surgical or nonoperative treatment for lumbar spinal stenosis? A randomized controlled trial. *Spine.* 2007;32(1):1-8.

49. Slatis P, Malmivaara A, Heliovaara M, et al. Long-term results of surgery for lumbar spinal stenosis: a randomised controlled trial. *European spine journal : official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society.* 2011;20(7):1174-1181.

50. Song SH, Ryu GH, Park JW, et al. The Effect and Safety of Steroid Injection in Lumbar Spinal Stenosis: With or Without Local Anesthetics. *Ann Rehabil Med.* 2016;40(1):14-20.

51. Stromqvist BH, Berg S, Gerdhem P, et al. X-stop versus decompressive surgery for lumbar neurogenic intermittent claudication: randomized controlled trial with 2-year follow-up. *Spine.* 2013;38(17):1436-1442.

52. Thome C, Zevgaridis D, Leheta O, et al. Outcome after less-invasive decompression of lumbar spinal stenosis: a randomized comparison of unilateral laminotomy, bilateral laminotomy, and laminectomy. *Journal of neurosurgery Spine.* 2005;3(2):129-141.

53. Watanabe K, Matsumoto M, Ikegami T, et al. Reduced postoperative wound pain after lumbar spinous process-splitting laminectomy for lumbar canal stenosis: a randomized controlled study. *Journal of neurosurgery Spine.* 2011;14(1):51-58.

54. Weinstein JN, Lurie JD, Tosteson TD, et al. Surgical versus nonsurgical treatment for lumbar degenerative spondylolisthesis. *The New England journal of medicine.* 2007;356(22):2257-2270.

55. Weinstein JN, Lurie JD, Tosteson TD, et al. Surgical compared with nonoperative treatment for lumbar degenerative spondylolisthesis. four-year results in the Spine Patient Outcomes Research Trial (SPORT) randomized and observational cohorts. *The Journal of bone and joint surgery American volume.* 2009;91(6):1295-1304.

56. Weinstein JN, Tosteson TD, Lurie JD, et al. Surgical versus nonsurgical therapy for lumbar spinal stenosis. *The New England journal of medicine.* 2008;358(8):794-810.

57. Yagi M, Okada E, Ninomiya K, Kihara M. Postoperative outcome after modified unilateral-approach microendoscopic midline decompression for degenerative spinal stenosis. *Journal of neurosurgery Spine.* 2009;10(4):293-299.

58. Hsu KY, Zucherman JF, Hartjen CA, et al. Quality of life of lumbar stenosis-treated patients in whom the X STOP interspinous device was implanted. *Journal of neurosurgery Spine.* 2006;5(6):500-507.

59. Zucherman JF, Hsu KY, Hartjen CA, et al. A prospective randomized multi-center study for the treatment of lumbar spinal stenosis with the X STOP interspinous implant: 1-year results. *European spine journal : official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society.* 2004;13(1):22-31.

60. Zucherman JF, Hsu KY, Hartjen CA, et al. A multicenter, prospective, randomized trial evaluating the X STOP interspinous process decompression system for the treatment of neurogenic intermittent claudication: two-year follow-up results. *Spine.* 2005;30(12):1351-1358.

CHAPTER FIVE

Association of Lumbar Spine Radiographic Changes With Severity of Back Pain-Related Disability Among Middle-aged, Community-Dwelling Women

Chapter Five has been published as:

Chen L.**,** Perera R.S., Radojčić M.R., Beckenkamp P.R., Ferreira P.H., Hart D.J., Spector T.D., Arden N.K., Ferreira M.L. Association of Lumbar Spine Radiographic Changes With Severity of Back Pain-Related Disability Among Middle-aged, Community-Dwelling Women. JAMA Network Open. 2021;4(5);e2110715.

Statement from co-authors confirming authorship contribution of the PhD candidate

The co-authors of the paper "Association of Lumbar Spine Radiographic Changes With Severity of Back Pain-Related Disability Among Middle-aged, Community-Dwelling Women" confirm that Lingxiao Chen has made the following contributions:

Conception and design of the research
Analysis and interpretation of the findings
Writing of the manuscript and critical appraisal of the content

In addition to the statements above, in cases where I am not the corresponding author of a published item, permission to include the published material has been granted by the corresponding author.

Lingxiao Chen Date: 21 July 2021

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.
Manuela Loureiro Ferreira Date: 21 July 2021

JAMA Network Open

Original Investigation | Imaging

Association of Lumbar Spine Radiographic Changes With Severity of Back Pain-Related Disability Among Middle-aged, Community-Dwelling Women

Lingxiao Chen, MBBS, MMed; Romain S. Perera, PhD; Maja R. Radojčić, PhD; Paula R. Beckenkamp, PhD; Paulo H. Ferreira, PhD; Deborah J. Hart, PhD; Tim D. Spector, PhD; Nigel K. Arden, PhD; Manuela L. Ferreira, PhD

Abstract

IMPORTANCE Previous studies, using mostly cross-sectional data, provide conflicting evidence of an association between lumbar spine radiographic changes and the severity of back pain-related disability. Such conflicting evidence may be associated with widely unnecessary diagnostic imaging of the lumbar spine.

OBJECTIVE To examine both cross-sectional and longitudinal associations between lumbar spine radiographic changes and the severity of back pain-related disability among middle-aged, community-dwelling women.

DESIGN, SETTING, AND PARTICIPANTS This population-based prospective cohort study used data from the Chingford 1000 Women Study. Analyses included data collected from year 6 (1994-1996; physical activity was measured), year 9 (1997-1999; treated as baseline), and year 15 (2003-2005), with a total length of follow-up for longitudinal analyses of 6 years. Data were analyzed from April 17 to November 3, 2020.

EXPOSURES Primary exposure was lumbar spine radiographic changes, defined using the Kellgren-Lawrence (K-L) grade. Secondary exposures were defined using presence of osteophytes and disc space narrowing. The composite score combined the number of lumbar spine segments with definite changes detected on radiographic images (ie, radiographic changes) (K-L grade ≥2, which means at least definite osteophyte and possible narrowing of disc space are present; osteophyte and disc space narrowing grade ≥1, which means at least mild or definite changes are present).

MAIN OUTCOMES AND MEASURES Self-reported back pain-related disability measured in years 9 and 15 assessed by the St Thomas disability questionnaire.

RESULTS Among 650 women (mean [SD] age, 61.3 [5.9] years) in cross-sectional analyses and 443 women (mean [SD] age, 60.6 [6.0] years) in longitudinal analyses, there was no evidence to support an association between higher number of lumbar segments with radiographic changes (K-L grade, osteophytes, and disc space narrowing) and more severe back pain-related disability (eg, cross-sectional analyses using the K-L grade; 1 segment vs 0 segment: adjusted odds ratio, 1.22 [95% CI, 0.76-1.96]). No interactions were found of an association between lumbar spine radiographic changes and the severity of back pain-specific disability with age, body mass index, or smoking status.

CONCLUSIONS AND RELEVANCE In this cohort of middle-aged, community-dwelling women, there was no evidence to support an association between a higher number of lumbar segments with radiographic changes (K-L grade, osteophytes, and disc space narrowing) and more severe back

(continued)

Key Points

Question Are lumbar spine radiographic changes associated with severity of back pain-related disability among middle-aged, community-dwelling women?

Findings In this population-based cohort study of women from the UK, there was no evidence to support an association between a higher number of lumbar segments with radiographic changes (Kellgren-Lawrence grade, osteophytes, and disc space narrowing) and more severe back pain-related disability in cross-sectional (650 women) or longitudinal (443 women) analyses.

Meaning The findings suggest that the changes detected on lumbar radiographs provide limited value for decision-making regarding back pain management in this population.

+ Supplemental content

Author affiliations and article information are listed at the end of this article.

Abstract (continued)

pain-related disability cross-sectionally or over time. These findings provide further evidence against routinely using diagnostic imaging of the lumbar spine.

JAMA Network Open. 2021;4(5):e2110715. doi:10.1001/jamanetworkopen.2021.10715

Introduction

Low back pain (LBP) is a highly prevalent condition in the general population worldwide and has been the leading cause of disability for nearly 3 decades, according to the Global Burden of Disease Study 2017.[1,2] Among all musculoskeletal problems, LBP is also the most common reason for patients to seek primary care.[3] Current guidelines for treatment of LBP do not recommend routinely using diagnostic imaging, except when patients either present with severe, progressive neurologic deficits or with signs or symptoms indicative of a serious or specific underlying condition (eg, fracture or cancer).[4-6] Nonetheless, diagnostic imaging is still widely used in clinical practice for LBP, with a recent meta-analysis indicating that more than 15% of patients in primary care and approximately 25% in emergency care receive a referral for simple imaging (mainly radiograph).[7] Moreover, nearly 10% of patients with LBP in primary care and emergency care also receive a referral for complex imaging (mainly computed tomography scan and magnetic resonance imaging).[7] There has also been a 53% relative increase in referrals for complex imaging from 1995 to 2017, with no change observed during that period for the rate of referrals for simple imaging.[7] Unnecessary diagnostic imaging not only wastes limited medical resources but is also associated with poorer health outcomes, such as iatrogenic disease from techniques that use ionizing radiation.[8] In addition, patients who undergo unnecessary diagnostic imaging might be labeled with a pseudodisease, which may be associated with unnecessary subsequent interventions that may have adverse effects.[9]

Possible explanations for the unwarranted prevalence of imaging referrals for LBP are (1) the patient's expectation that imaging results could provide valuable information on the cause and, consequently, the appropriate management of her or his condition and (2) the clinician's desire to reassure the patient of the absence of any underlying pathologic condition.[10-12] Previous studies have confirmed that imaging does not improve clinical outcomes for patients with LBP.[9,13] However, the definition of normal or abnormal imaging diagnostic findings is still debatable. Currently, the presence of osteophytes and disc space narrowing are the most frequent changes detected on radiographs (hereafter referred to as radiographic changes) that may be indicative of spinal pathologic conditions, and the Kellgren-Lawrence (K-L) grade is the tool commonly used to assess the severity of osteoarthritis.[14] A study including elderly women who lived in rural South Korea showed a positive association of the presence of osteophytes (grade ≥2), disc space narrowing (grade ≥2), and K-L grade (grade ≥2) with the severity of disability, measured by a validated Korean version of the Oswestry Disability Index.[15] However, a study conducted in Sri Lanka including patients with LBP concluded that neither disc space narrowing nor the presence of osteophytes was associated with the severity of disability (also measured with the Modified Oswestry Disability Index).[16] Both studies are cross-sectional and failed to adjust their analyses for important confounders, including smoking status, level of participation in physical activity, and medication use. Past studies have also failed to identify whether the number of affected lumbar segments is associated with the severity of back pain-related disability. Therefore, the role of radiographic findings as a potential prognostic factor of the clinical course of LBP is still unclear and needs to be fully explored in population-based cohort studies. Previous studies of radiographic changes in knee osteoarthritis have indicated that the presence of osteophytes may be used to diagnose the condition and that the presence of joint space narrowing may be used to assess both the diagnosis and the progression of osteoarthritis.[17,18] This is still to be elucidated among patients with LBP.

The aim of this study was to examine both cross-sectional and longitudinal associations between lumbar radiographic changes and the severity of back pain–related disability among middle-aged, community-dwelling women using composite scores that combined the number of segments and type of changes in terms of K-L grade, disc space narrowing, and osteophytes. We hypothesized that a higher number of segments with lumbar radiographic changes would be associated with more severe back pain–related disability.

Methods

Study Design, Data Sources, and Study Population

From an age and sex register of a large practice of more than 11 000 patients in Chingford in east London, UK, all 1353 women in the age range of 45 to 64 years were invited to participate in a population study assessing musculoskeletal diseases. A total of 1003 women were examined between 1989 and 1991 (year 1; baseline visit for original cohort); 6 died, 66 had moved away, and 278 refused to participate or did not respond. All the women lived within 8 km (5 miles) of the general practice, and 98% of the women were white. Women from this general practice are similar to the UK general population in terms of weight, height, and body mass index (BMI). In the data analyses, and given their availability, we included data on physical activity collected in year 6 (1994-1996 [ie, prebaseline]) and imaging data, all other covariates, and the outcome for cross-sectional analyses collected in year 9 (1997-1999 [ie, baseline for our study]). The outcome for longitudinal analyses was obtained in year 15 (2003-2005). We followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline.[19,20] The Waltham Forest and Redbridge local research ethics committee approved the study, and all participants provided written informed consent to participate in the study.

Exposures

Lateral lumbar spine radiographs at year 9 were taken by 1 radiographer, centered on the L3 vertebra, with the participants in the left lateral recumbent position. A single trained observer (a rheumatologist) blinded to patient identity and chronologic order read all of the radiographs. Within-observer variation was assessed by test-retest analysis of 40 randomly selected radiographs from the study. Good within-observer reproducibility (κ = 0.78-0.89) was found.[21] At each lumbar spine segment (L1-L2, L2-L3, L3-L4, and L4-L5), disc space narrowing and osteophytes (both anterior and posterior) were assessed through the semiquantitative method reported by Lane et al,[22] with grade 0 corresponding to normal, grade 1 to mild narrowing and osteophytes, grade 2 to moderate narrowing and osteophytes, and grade 3 to severe narrowing and osteophytes. The Kellgren-Lawrence (K-L) grade was summarized as grade 0 indicating normal; grade 1 indicating doubtful narrowing of disc space and possible osteophytic lipping; grade 2 indicating definite osteophyte and possible narrowing of disc space; grade 3 indicating moderate multiple osteophytes, definite narrowing of disc space, some sclerosis, and possible deformity of bone contour; and grade 4 indicating large osteophytes, marked narrowing of disc space, severe sclerosis, and definite deformity of bone contour.

Considering that the number of lumbar spine segments with radiographic changes detected and the various types of radiographic changes might be associated with the results, we generated 3 composite scores: a K-L grade–based score, an osteophyte grade–based score, and a disc space narrowing grade–based score; at each segment, a binary exposure variable of 1 (K-L grade ≥2, which means at least definite osteophyte and possible narrowing of disc space are present; disc space narrowing and osteophyte grade ≥1, which means at least mild or definite changes are present) vs 0 (K-L grade 0 or 1; disc space narrowing and osteophyte grade 0) was used. The composite score was then calculated as the final L1-L2 score + L2-L3 score + L3-L4 score + L4-L5 score, with values ranging from 0 to 4 (where 0 indicates no lumbar spine segments with radiographic changes detected and 4 indicates 4 lumbar spine segments with radiographic changes detected). The K-L grade–based

score was defined as the primary exposure. Osteophyte and disc space narrowing grade–based scores were set as secondary exposures.

Outcomes

Back pain–related disability was assessed at year 9 and year 15 using a back pain questionnaire (St Thomas disability questionnaire), which correlated well with the Oswestry Disability Questionnaire ($r = 0.77$; $P < .001$).[23] The outcome was defined by questions at 2 levels. At the first level, women were asked whether they had any back pain for at least 1 day at any time in the last 12 months. At the second level, those who answered yes to the first-level question were asked 8 questions related to the disability due to back pain (corresponding to the previous year's status): walking around the house; standing for 15 minutes; getting up from a low chair; getting out of a bath; getting in and out of a car; going up and down stairs; putting on socks, stockings, or tights; and cutting toenails. Each question was summarized as grade 0 indicating no difficulty, grade 1 indicating difficult but possible, and grade 2 indicating impossible. We built a composite score based on the aforementioned 8 questions; values ranged from 0 to 16, with higher values corresponding to more severe disability. We assumed the composite score as 0 if women answered "no" to the first-level question. In case of missing data for any of the 8 questions, we kept the data if the women responded to at least 6 questions and calculated the composite score as [(total score)/(number of questions answered)] × 8.

Covariates

Causal diagram through DAGitty, version 3.0[24] was used to choose the minimal sufficient adjustment sets for estimating the total association of the exposure with the outcome.[25] Age, BMI, smoking status, back pain status, bisphosphonate use, and physical activity were included in the final model (details in eFigure 1 in the Supplement). All covariates, except physical activity, which was measured in year 6, were measured in year 9.

Statistical Analysis

Data were analyzed from April 17 to November 3, 2020. Owing to the skewed distribution of back pain–related disability (eFigure 2 in the Supplement), ordinal logistic regression, which holds a proportional odds assumption, was performed.[26] Considering that physical activity was measured at a different time point (ie, year 6) compared with other covariates (ie, year 9) and with potential measurement error, we established a stepped modeling framework: step 1, unadjusted analyses; step 2, analyses adjusted for age, BMI, back pain status, bisphosphonate use status, and smoking status (additionally adjusted for year 9 back pain–related disability for the longitudinal analysis); and step 3, analyses further adjusted for physical activity.

Separate analyses were conducted for cross-sectional and longitudinal data. For the longitudinal analyses, data on lumbar spine radiographic changes collected in year 9 were treated as the exposure, and back pain–related disability data collected in year 15 were treated as the outcome. In addition to the confounders mentioned, data on back pain–related disability collected in year 9 were included in the longitudinal analysis as a strong prognostic factor to adjust.[27] Based on the recommendation from *Modern Epidemiology*,[28] the exposures were modeled as unordered categorical variables and trend test.[26]

The proportion of missing data in each covariate is provided in **Table 1**. Missing data were handled through multiple imputation, which holds a missing-at-random assumption.[26] The assumption was graphically tested (eFigure 3 in the Supplement). No additional variables were used; all covariates in the minimal sufficient adjustment sets were used. Flexible additive models with 10 imputed data sets were used.[29] We did not impute data for the exposure variables. The relative risk was presented as adjusted proportional odds ratios (ORs) with 95% CIs. Extensive sensitivity analyses were performed (eAppendix 1 in the Supplement). All statistical analyses were performed with rms, Hmisc, and tidyverse packages in R, version 3.6.2 (R Group for Statistical Computing). Details of the statistical methods are provided in eAppendix 2 in the Supplement.

Table 1. Baseline Characteristics of Study Participants

Characteristic	Participants, No. (%)					
	No. of segments[a]					
	0	1	2	3	4	Whole cohort
Cross-sectional (n = 650)						
No.	154	142	140	118	96	650
Age, mean (SD), y	59.7 (5.5)	60.3 (5.7)	60.8 (5.8)	62.8 (5.7)	64.5 (5.8)	61.3 (5.9)
Missing	1 (0.6)	2 (1.4)	4 (2.9)	2 (1.7)	1 (1.0)	10 (1.5)
Smoking status						
Never	105 (68.2)	84 (59.2)	89 (63.6)	71 (60.2)	61 (63.5)	410 (63.1)
Current	18 (11.7)	25 (17.6)	20 (14.3)	20 (16.9)	15 (15.6)	98 (15.1)
Ex-smoker	28 (18.2)	31 (21.8)	25 (17.9)	23 (19.5)	18 (18.8)	125 (19.2)
Missing	3 (1.9)	2 (1.4)	6 (4.3)	4 (3.4)	2 (2.1)	17 (2.6)
BMI, mean (SD)	26.6 (4.2)	26.8 (5.1)	27.2 (5.0)	26.9 (4.4)	27.6 (5.4)	27.0 (4.8)
Missing	1 (0.6)	3 (2.1)	4 (2.9)	4 (3.4)	1 (1.0)	13 (2.0)
Back pain status						
Yes	51 (33.1)	46 (32.4)	39 (27.9)	42 (35.6)	32 (33.3)	210 (32.3)
No	103 (66.9)	96 (67.6)	101 (72.1)	76 (64.4)	64 (66.7)	440 (67.7)
Bisphosphonate use status						
Yes	5 (3.2)	3 (2.1)	4 (2.9)	1 (0.8)	1 (1.0)	14 (2.2)
No	68 (44.2)	62 (43.7)	71 (50.7)	55 (46.6)	44 (45.8)	300 (46.2)
Missing	81 (52.6)	77 (54.2)	65 (46.4)	62 (52.5)	51 (53.1)	336 (51.7)
Physical activity						
Walking, km/wk						
<0.8	14 (9.1)	6 (4.2)	12 (8.6)	7 (5.9)	13 (13.5)	52 (8.0)
0.8 to <8.1	75 (48.7)	75 (52.8)	60 (42.9)	70 (59.3)	43 (44.8)	323 (49.7)
8.1 to <16.1	42 (27.3)	36 (25.4)	41 (29.3)	30 (25.4)	17 (17.7)	166 (25.5)
≥16.1	17 (11.0)	19 (13.4)	21 (15.0)	5 (4.2)	16 (16.7)	78 (12.0)
Missing	6 (3.9)	6 (4.2)	6 (4.3)	6 (5.1)	7 (7.3)	31 (4.8)
Job						
Sedentary	8 (5.2)	8 (5.6)	4 (2.9)	4 (3.4)	0	24 (3.7)
Sedentary plus occasional exercise	23 (14.9)	14 (9.9)	16 (11.4)	13 (11.0)	8 (8.3)	74 (11.4)
0.5 Sedentary plus 0.5 active (or active housework [eg, daily dusting or vacuuming])	88 (57.1)	77 (54.2)	87 (62.1)	73 (61.9)	59 (61.5)	384 (59.1)
Predominantly manual, active all day	27 (17.5)	30 (21.1)	22 (15.7)	17 (14.4)	15 (15.6)	111 (17.1)
Missing	8 (5.2)	13 (9.2)	11 (7.9)	11 (9.3)	14 (14.6)	57 (8.8)
Sport						
None	80 (51.9)	78 (54.9)	89 (63.6)	70 (59.3)	60 (62.5)	377 (58.0)
Golf, bowling, badminton, cycling, or swimming, 1 h/wk	22 (14.3)	26 (18.3)	15 (10.7)	15 (12.7)	8 (8.3)	86 (13.2)
2 h/wk of Golf, bowling, badminton, cycling, or swimming or 1 h of staying fit, aerobics, or squash	30 (19.5)	23 (16.2)	22 (15.7)	19 (16.1)	11 (11.5)	105 (16.2)
≥2 h/wk of Staying fit, aerobics, or squash	17 (11.0)	9 (6.3)	8 (5.7)	8 (6.8)	11 (11.5)	53 (8.2)
Missing	5 (3.2)	6 (4.2)	6 (4.3)	6 (5.1)	6 (6.3)	29 (4.5)
Disability, year 9, median (IQR)	0 (0.0-5.8)	0 (0.0-6.0)	0 (0.0-3.3)	0 (0.0-4.9)	0 (0.0-3.3)	0 (0.0-5.0)
Longitudinal (n = 443)						
No.	112	100	97	76	58	443
Age, mean (SD), y	59.1 (5.3)	59.5 (5.7)	60.5 (6.3)	62.6 (6.1)	62.8 (5.9)	60.6 (6.0)
Missing	1 (0.9)	2 (2.0)	3 (3.1)	0	0	6 (1.4)

(continued)

Table 1. Baseline Characteristics of Study Participants (continued)

Characteristic	Participants, No. (%)					
	No. of segments[a]					
	0	1	2	3	4	Whole cohort
Smoking status						
Never	77 (68.8)	63 (63.0)	64 (66.0)	44 (57.9)	38 (65.5)	286 (64.6)
Current	13 (11.6)	14 (14.0)	11 (11.3)	16 (21.1)	9 (15.5)	63 (14.2)
Ex-smoker	19 (17.0)	21 (21.0)	18 (18.6)	15 (19.7)	11 (19.0)	84 (19.0)
Missing	3 (2.7)	2 (2.0)	4 (4.1)	1 (1.3)	0	10 (2.3)
BMI, mean (SD)	26.4 (4.3)	26.7 (4.9)	27.2 (5.2)	26.5 (3.7)	28.3 (5.3)	26.9 (4.7)
Missing	1 (0.9)	3 (3.0)	3 (3.1)	2 (2.6)	0	9 (2.0)
Back pain status						
Yes	40 (35.7)	32 (32.0)	24 (24.7)	26 (34.2)	23 (39.7)	145 (32.7)
No	72 (64.3)	68 (68.0)	73 (75.3)	50 (65.8)	35 (60.3)	298 (67.3)
Bisphosphonate use status						
Yes	4 (3.6)	3 (3.0)	2 (2.1)	1 (1.3)	1 (1.7)	11 (2.5)
No	43 (38.4)	43 (43.0)	55 (56.7)	35 (46.1)	26 (44.8)	202 (45.6)
Missing	65 (58.0)	54 (54.0)	40 (41.2)	40 (52.6)	31 (53.4)	230 (51.9)
Physical activity						
Walking, km/wk						
<0.8	11 (9.8)	5 (5.0)	12 (12.4)	5 (6.6)	9 (15.5)	42 (9.5)
0.8 to <8.1	54 (48.2)	50 (50.0)	42 (43.3)	46 (60.5)	29 (50.0)	221 (49.9)
8.1 to <16.1	31 (27.7)	27 (27.0)	22 (22.7)	20 (26.3)	8 (13.8)	108 (24.4)
≥16.1	10 (8.9)	16 (16.0)	17 (17.5)	3 (3.9)	10 (17.2)	56 (12.6)
Missing	6 (5.4)	2 (2.0)	4 (4.1)	2 (2.6)	2 (3.5)	16 (3.6)
Job						
Sedentary	5 (4.5)	5 (5.0)	4 (4.1)	3 (3.9)	0	17 (3.8)
Sedentary plus occasional exercise	15 (13.4)	13 (13.0)	10 (10.3)	9 (11.8)	6 (10.3)	53 (12.0)
0.5 Sedentary plus 0.5 active (or active housework [eg, daily dusting or vacuuming])	65 (58.0)	55 (55.0)	62 (63.9)	49 (64.5)	37 (63.8)	268 (60.5)
Predominantly manual, active all day	21 (18.8)	24 (24.0)	14 (14.4)	11 (14.5)	10 (17.2)	80 (18.1)
Missing	6 (5.4)	3 (3.0)	7 (7.2)	4 (5.3)	5 (8.6)	25 (5.6)
Sport						
None	60 (53.6)	52 (52.0)	57 (58.8)	45 (59.2)	38 (65.5)	252 (56.9)
Golf, bowling, badminton, cycling, or swimming, 1 h/wk	14 (12.5)	21 (21.0)	15 (15.5)	8 (10.5)	4 (6.9)	62 (14.0)
2 h/wk of Previous or 1 h of staying fit, aerobics, or squash	22 (19.6)	19 (19.0)	16 (16.5)	17 (22.4)	6 (10.3)	80 (18.1)
≥2 h/wk of Staying fit, aerobics, or squash	11 (9.8)	6 (6.0)	4 (4.1)	4 (5.3)	8 (13.8)	33 (7.4)
Missing	5 (4.5)	2 (2.0)	5 (5.2)	2 (2.6)	2 (3.4)	16 (3.6)
Disability, year 15, median (IQR)	0 (0.0-5.0)	0 (0.0-5.1)	0 (0.0-4.0)	0 (0.0-2.5)	0 (0.0-4.8)	0 (0.0-4.8)

Abbreviations: BMI, body mass index (calculated as weight in kilograms divided by height in meters squared); IQR, interquartile range.

[a] Number of segments of lumbar spine radiographic changes (based on Kellgren-Lawrence grade).

Results

Participant Characteristics

A total of 650 women (mean [SD] age, 61.3 [5.9] years) were included in cross-sectional analyses, and a total of 443 women (mean [SD] age, 60.6 [6.0] years) were included in longitudinal analyses (Table 1; **Figure**). Most study participants were classified as either never smokers or ex-smokers (Table 1). The median score of back pain–related disability was 0 (interquartile range, 0.0-5.0 in cross-sectional analyses and 0.0-4.8 in longitudinal analyses) in both cross-sectional and longitudinal analyses. The distribution of each lumbar spine radiographic change at each lumbar spine segment

is listed in eTable 1 in the Supplement. Redundancy analyses were performed to assess whether 1 exposure could be estimated from any 2 other exposures, at each lumbar spine segment.[26] No exposure was redundant (eTable 2 in the Supplement).

K-L Grade–Based Score

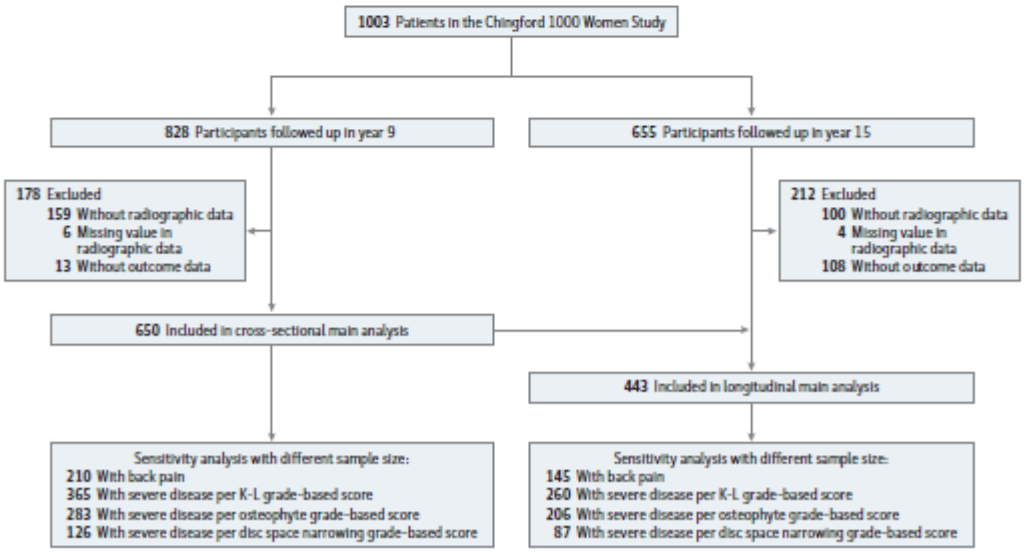
Using a multivariable ordinal logistic regression model, we found that women who had 1 or more segments with lumbar spine radiographic K-L grade–based changes were not statistically more likely to report more disability compared with women with no observed changes in both cross-sectional (eg, 1 segment vs 0 segments; step 2 model; OR, 1.22 [95% CI, 0.76-1.96]) and longitudinal analyses (**Table 2**). When further adjustment was made for physical activity, the results were similar (OR, 1.19 [95% CI, 0.73-1.93]). No evidence was found to support a linear or nonlinear trend between number of segments with lumbar spine radiographic changes and the severity of back pain–related disability.

Osteophyte Grade–Based Score and Disc Space Narrowing Grade–Based Score

For osteophyte grade–based score, no statistically significant association was found between the number of lumbar segments with radiographic changes and the severity of back pain–related disability in both cross-sectional (eg, 1 segment vs 0 segment; step 2 model; OR, 0.83 [95% CI, 0.57-1.22]) and longitudinal analyses (**Table 3**). Similar results were observed when further adjustments were made in the models to account for participation in physical activity in both cross-sectional and longitudinal analyses. In the longitudinal analysis, a greater number of affected segments were linearly associated with less severe back pain–related disability (step 2 model).

For the disc space narrowing grade–based score, no statistically significant association was observed between the number of lumbar segments with radiographic changes and the severity of back pain–related disability in both cross-sectional (eg, 1 segment vs 0 segment; step 2 model; OR, 1.43 [95% CI, 0.78-2.61]) and longitudinal analyses (**Table 4**). Results were similar when we further adjusted for physical activity (OR, 1.41 [95% CI, 0.77-2.60]). For the cross-sectional analyses, a higher

Figure. Flowchart of Study Participants



A total of 95.0% women (421 of 443) in the longitudinal analysis are also in the cross-sectional analysis. Besides all data mentioned in the flowchart, we also collected the data from year 6 (1994-1996) because it was when 1 necessary covariate (ie, physical activity) was measured. K-L indicates Kellgren-Lawrence.

number of segments of lumbar spine radiographic characteristics were nonlinearly associated with less severe back pain–related disability (step 2 model).

Exploratory and Sensitivity Analyses

We did not find interactions with age, BMI, or smoking status (eTable 3 in the Supplement) between lumbar spine radiographic changes and the severity of back pain–related disability. Overall, our results remained similar under extensive sensitivity analyses (eAppendices 3-8 and eTables 4-25 in the Supplement). All E-values are listed in eTable 26 in the Supplement.

Discussion

Key Results

In this cohort of middle-aged women from Chingford in east London, UK, no evidence was found to support an association between a higher number of segments with lumbar radiographic changes (K-L grade, osteophyte, and disc space narrowing) and more severe back pain–related disability. Our

Table 2. Association Between K-L Grade–Based Score and Severity of Back Pain–Related Disability

	K-L grade-based score, OR (95% CI)					*P* value for trend	
Variable	0 Segments	1 Segment	2 Segments	3 Segments	4 Segments	Linear model	Nonlinear model
Cross-sectional, year 9 (n = 650)							
Women, No. (%)	154 (23.7)	142 (21.8)	140 (21.5)	118 (18.2)	96 (14.8)	NA	NA
Unadjusted	1 [Reference]	1.19 (0.75-1.89)	0.84 (0.52-1.36)	0.90 (0.54-1.49)	0.81 (0.47-1.38)	.24	.78
Multivariable adjusted[a]	1 [Reference]	1.22 (0.76-1.96)	0.84 (0.51-1.38)	0.92 (0.54-1.56)	0.89 (0.50-1.57)	.39	.91
Further adjusted for physical activity	1 [Reference]	1.19 (0.73-1.93)	0.81 (0.49-1.35)	0.87 (0.51-1.49)	0.85 (0.48-1.50)	.30	.96
Longitudinal, year 15 (n = 443)							
Women, No. (%)	112 (25.3)	100 (22.6)	97 (21.9)	76 (17.2)	58 (13.1)	NA	NA
Unadjusted	1 [Reference]	1.22 (0.70-2.11)	0.95 (0.53-1.68)	0.77 (0.41-1.44)	1.05 (0.55-2.02)	.57	.92
Multivariable adjusted[b]	1 [Reference]	1.06 (0.57-1.96)	0.94 (0.50-1.76)	0.69 (0.34-1.38)	0.83 (0.40-1.72)	.20	.79
Further adjusted for physical activity	1 [Reference]	1.08 (0.58-2.02)	0.91 (0.48-1.72)	0.67 (0.33-1.38)	0.80 (0.38-1.68)	.17	.80

Abbreviations: K-L, Kellgren-Lawrence; NA, not applicable; OR, odds ratio.

[a] Adjusted for age, body mass index, smoking status, back pain status, and bisphosphonate use.

[b] Adjusted for age, body mass index, smoking status, back pain status, bisphosphonate use, and year 9 back pain–related disability.

Table 3. Association Between Osteophyte Grade–Based Score and the Severity of Back Pain–Related Disability

	Osteophyte grade-based score, OR (95% CI)					*P* value for trend	
Variable	0 Segments	1 Segment	2 Segments	3 Segments	4 Segments	Linear model	Nonlinear model
Cross-sectional, year 9 (n = 650)							
Women, No. (%)	258 (39.7)	226 (34.8)	102 (15.7)	44 (6.8)	20 (3.1)	NA	NA
Unadjusted	1 [Reference]	0.80 (0.55-1.16)	0.81 (0.50-1.32)	0.53 (0.25-1.12)	0.81 (0.31-2.14)	.12	.59
Multivariable adjusted[a]	1 [Reference]	0.83 (0.57-1.22)	0.78 (0.47-1.30)	0.58 (0.27-1.26)	1.03 (0.37-2.85)	.25	.42
Further adjusted for physical activity	1 [Reference]	0.82 (0.56-1.21)	0.79 (0.47-1.32)	0.60 (0.28-1.31)	0.98 (0.35-2.73)	.26	.46
Longitudinal, year 15 (n = 443)							
Women, No. (%)	192 (43.3)	157 (35.4)	67 (15.1)	19 (4.3)	8 (1.8)	NA	NA
Unadjusted	1 [Reference]	0.71 (0.46-1.10)	0.60 (0.33-1.08)	0.41 (0.13-1.27)	0.49 (0.10-2.40)	.02	.65
Multivariable adjusted[b]	1 [Reference]	0.76 (0.47-1.24)	0.53 (0.28-1.02)	0.49 (0.14-1.70)	0.31 (0.06-1.72)	.01	.75
Further adjusted for physical activity	1 [Reference]	0.76 (0.46-1.24)	0.52 (0.27-1.03)	0.50 (0.14-1.74)	0.33 (0.06-1.79)	.01	.68

Abbreviations: NA, not applicable; OR, odds ratio.

[a] Adjusted for age, body mass index, smoking status, back pain status, and bisphosphonate use.

[b] Adjusted for age, body mass index, smoking status, back pain status, bisphosphonate use, and year 9 back pain–related disability.

results remained unchanged after including potential interactions with important confounders, such as age, BMI, and smoking status, and after extensive sensitivity analyses.

Comparison With Previous Studies

For K-L results, our findings contradict those of Lee et al,[15] who found that K-L grades were significantly associated with the Oswestry Disability Index. The main reason for such a discrepancy in results may be the design features of the study by Lee et al,[15] which only included cross-sectional analyses with insufficient adjustment for important confounders (eg, physical activity, smoking status, and BMI). For osteophyte results, the results from the cross-sectional analyses are consistent with a previous cross-sectional study by Perera et al,[16] who identified that the presence of osteophytes on radiographs was not associated with physical disability measured with the Oswestry Disability Index. However, results from our longitudinal analyses indicated that a greater number of affected segments were linearly associated with less severe back pain–related disability. One possible explanation is the biomechanical stability provided by spinal osteophytes, which have been proven to increase spinal resistance in compression.[30] Another explanation is that the results simply reflect large numbers of analyses completed, which need future studies to verify. For disc space narrowing results, our results are similar to those of Perera et al[16] but different than the findings of Lee et al.[15] Such differences could also be associated with the methodological limitations already described in the study conducted by Lee et al.[15] Overall, the prevalence of lumbar spine radiographic findings in our study is similar to that of previous studies that indicated that many imaging-based spinal radiographic changes are likely part of normal, asymptomatic aging.[31]

Strengths and Limitations

Our study has several strengths. To our knowledge, this is the first study to create a composite score that reflects the overall association of lumbar spine radiographic changes (ie, number of affected segments and severity of changes) with the severity of back pain–related disability. We used population-based data that contain a long-term follow-up with good recruitment and retention rates, and multiple potential confounders were measured. We also overcame some methodological limitations from previous studies; we included a cohort study design and incorporated a systematic way to select and control confounders, we have repeated measures of back pain–related disability that allow us to adjust for baseline disability, we assessed the potential interaction term, and we performed extensive sensitivity analyses to evaluate the robustness of the results.

Limitations also need to be considered. First, the Chingford 1000 Women Study included middle-aged women in a specific area of the UK. We must exercise caution when generalizing the

Table 4. Association Between Disc Space Narrowing Grade–Based Score and the Severity of Back Pain–Related Disability

	Disc space narrowing grade-based score, OR (95% CI)					*P* value for trend	
Variable	0 Segments	1 Segment	2 Segments	3 Segments	4 Segments	Linear model	Nonlinear model
Cross-sectional, year 9 (n = 650)							
Women, No. (%)	100 (15.4)	107 (16.5)	147 (22.6)	131 (20.2)	165 (25.4)	NA	NA
Unadjusted	1 [Reference]	1.33 (0.74-2.37)	1.30 (0.75-2.24)	1.33 (0.77-2.31)	0.98 (0.57-1.69)	.75	.11
Multivariable adjusted[a]	1 [Reference]	1.43 (0.78-2.61)	1.56 (0.88-2.76)	1.44 (0.81-2.57)	1.07 (0.60-1.92)	.94	.04
Further adjusted for physical activity	1 [Reference]	1.41 (0.77-2.60)	1.56 (0.87-2.77)	1.45 (0.81-2.59)	1.04 (0.57-1.87)	.86	.03
Longitudinal, year 15 (n = 443)							
Women, No. (%)	70 (15.8)	84 (19.0)	102 (23.0)	88 (19.9)	99 (22.3)	NA	NA
Unadjusted	1 [Reference]	0.91 (0.47-1.75)	0.78 (0.41-1.49)	1.07 (0.56-2.05)	1.24 (0.67-2.29)	.34	.25
Multivariable adjusted[b]	1 [Reference]	0.72 (0.34-1.53)	0.74 (0.36-1.52)	1.06 (0.52-2.20)	1.26 (0.62-2.57)	.18	.18
Further adjusted for physical activity	1 [Reference]	0.68 (0.31-1.48)	0.67 (0.31-1.43)	1.12 (0.53-2.34)	1.33 (0.63-2.80)	.13	.12

Abbreviations: NA, not applicable; OR, odds ratio.

[a] Adjusted for age, body mass index, smoking status, back pain status, and bisphosphonate use.

[b] Adjusted for age, body mass index, smoking status, back pain status, bisphosphonate use, and year 9 back pain–related disability.

results to men, other age groups, other racial/ethnic groups, or other countries. Second, as with most studies, there is the potential for residual confounding (eg, participation in physical activity was measured 3 years before baseline). Third, the labeling of the images may have introduced potential bias in our results, given that there was only 1 observer and that lumbar spine levels were decided by the clinical experience. Fourth, owing to data unavailability, we could not establish whether there was any association between other radiologic changes, including spondylolisthesis or vertebral body height (ie, osteoporotic fractures), and severity of functional limitation. Fifth, although our outcome correlated well with the Oswestry Disability Questionnaire, it lacked strict validation. Sixth, although we aimed to focus on LBP–related disability, we only had a back pain variable, which might be slightly different from LBP.

Implications for Practice and Research

Clinicians may use the results of this study to educate patients and their colleagues that lumbar radiographic findings cannot provide prognostic information on back pain–related disability, further adding to the evidence supporting the urge to reduce unnecessary imaging referrals. Future studies should include participants of both sexes and larger sample sizes and should include multiple centers to increase external validity. The association between the findings of complex imaging (eg, computed tomography scans, magnetic resonance imaging, or nuclear bone scans) and symptom severity in people with LBP needs to be further explored, considering the increasing use of such imaging.

Conclusions

In this cohort of middle-aged, community-dwelling women, there was no evidence to support an association between a higher number of lumbar segments with radiographic changes (K-L grade, osteophytes, and disc space narrowing) and more severe back pain–related disability cross-sectionally or over time. The findings suggest that the changes detected on lumbar radiographs provide limited value for decision-making regarding back pain management in this population.

ARTICLE INFORMATION

Accepted for Publication: March 24, 2021.

Published: May 20, 2021. doi:10.1001/jamanetworkopen.2021.10715

Corresponding Author: Lingxiao Chen, MBBS, MMed, Institute of Bone and Joint Research, The Kolling Institute, Northern Clinical School, Faculty of Medicine and Health, University of Sydney, Sydney, NSW 2064, Australia (lche4036@uni.sydney.edu.au).

Author Affiliations: Institute of Bone and Joint Research, The Kolling Institute, Northern Clinical School, Faculty of Medicine and Health, University of Sydney, Sydney, New South Wales, Australia (Chen, M. L. Ferreira); Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, United Kingdom (Perera, Radojčić, Arden); Department of Allied Health Sciences, Faculty of Medicine, University of Colombo, Colombo, Sri Lanka (Perera); Centre for Sport, Exercise and Osteoarthritis Research Versus Arthritis, University of Oxford, Oxford, United Kingdom (Radojčić); School of Health Sciences, Faculty of Medicine and Health, University of Sydney, Sydney, New South Wales, Australia (Beckenkamp, P. H. Ferreira); Department of Twin Research and Genetic Epidemiology, King's College London, London, United Kingdom (Hart, Spector); Medical Research Council Environmental Epidemiology Unit, University of Southampton, Southampton, United Kingdom (Arden).

Author Contributions: Drs Chen and Perera had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Drs Arden and M. L. Ferreira contributed equally as co-senior authors.

Concept and design: Chen, Perera, Radojčić, Beckenkamp, P.H. Ferreira, Spector, Arden, M. L. Ferreira.

Acquisition, analysis, or interpretation of data: Chen, Perera, Beckenkamp, Hart, Arden, M. L. Ferreira.

Drafting of the manuscript: Chen, Beckenkamp, Spector, Arden.

Critical revision of the manuscript for important intellectual content: Chen, Perera, Radojčić, Beckenkamp, P. H. Ferreira, Hart, Arden, M. L. Ferreira.

Statistical analysis: Chen, Spector.

Obtained funding: Hart.

Administrative, technical, or material support: Chen, Perera, Hart.

Supervision: Beckenkamp, P. H. Ferreira, Arden, M. L. Ferreira.

Conflict of Interest Disclosures: Dr Spector reported serving as a scientific consultant for Zoe Global Ltd. Dr Arden reported receiving personal fees from Pfizer/Lilly and Bristows LLP and grants from Merck outside the submitted work. No other disclosures were reported.

Funding/Support: Arthritis Research UK (now Versus Arthritis) and the Oxford NIHR Musculoskeletal Biomedical Research Unit provided funding support to the study.

Role of the Funder/Sponsor: The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Additional Contributions: We would like to thank all the participants of the Chingford 1000 Women Study; Alan Hakim, MA, Rheumatology, Platinum Medical Centre, The Wellington Hospital, research coordination; Maxine Daniels, BSc(Hons), University of Oxford, research coordination; and Alison Turner, BSc, NIHR Musculoskeletal Biomedical Research Unit, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, development of data management system, for their time and dedication. We would like to thank Julie Damnjanović, BSc, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, for her valued support to the team and the study data management. None of the individuals named here were compensated for their contributions.

REFERENCES

1. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet*. 2018;392(10159):1789-1858. doi:10.1016/S0140-6736(18)32279-7

2. Hartvigsen J, Hancock MJ, Kongsted A, et al; Lancet Low Back Pain Series Working Group. What low back pain is and why we need to pay attention. *Lancet*. 2018;391(10137):2356-2367. doi:10.1016/S0140-6736(18)30480-X

3. Jordan KP, Kadam UT, Hayward R, Porcheret M, Young C, Croft P. Annual consultation prevalence of regional musculoskeletal problems in primary care: an observational study. *BMC Musculoskelet Disord*. 2010;11:144. doi:10.1186/1471-2474-11-144

4. Chou R, Qaseem A, Owens DK, Shekelle P; Clinical Guidelines Committee of the American College of Physicians. Diagnostic imaging for low back pain: advice for high-value health care from the American College of Physicians. *Ann Intern Med*. 2011;154(3):181-189. doi:10.7326/0003-4819-154-3-201102010-00008

5. Foster NE, Anema JR, Cherkin D, et al; Lancet Low Back Pain Series Working Group. Prevention and treatment of low back pain: evidence, challenges, and promising directions. *Lancet*. 2018;391(10137):2368-2383. doi:10.1016/S0140-6736(18)30489-6

6. National Institute of Health and Care Excellence (NICE); National Guideline Centre. *Low Back Pain and Sciatica in Over 16s: Assessment And Management*. NICE Guideline No. 59. National Institute for Health and Care Excellence; 2016.

7. Downie A, Hancock M, Jenkins H, et al. How common is imaging for low back pain in primary and emergency care? systematic review and meta-analysis of over 4 million imaging requests across 21 years. *Br J Sports Med*. 2020;54(11):642-651. doi:10.1136/bjsports-2018-100087

8. Wall BF, Kendall GM, Edwards AA, Bouffler S, Muirhead CR, Meara JR. What are the risks from medical X-rays and other low dose radiation? *Br J Radiol*. 2006;79(940):285-294. doi:10.1259/bjr/55733882

9. Chou R, Fu R, Carrino JA, Deyo RA. Imaging strategies for low-back pain: systematic review and meta-analysis. *Lancet*. 2009;373(9662):463-472. doi:10.1016/S0140-6736(09)60172-0

10. Rhodes LA, McPhillips-Tangum CA, Markham C, Klenk R. The power of the visible: the meaning of diagnostic tests in chronic back pain. . 1999;48(9):1189-1203. doi:10.1016/s0277-9536(98)00418-3

11. Schers H, Wensing M, Huijsmans Z, van Tulder M, Grol R. Implementation barriers for general practice guidelines on low back pain: a qualitative study. *Spine (Phila Pa 1976)*. 2001;26(15):E348-E353. doi:10.1097/00007632-200108010-00013

12. Taylor S, Bishop A. Patient and public beliefs about the role of imaging in the management of non-specific low back pain: a scoping review. *Physiotherapy*. 2020;107:224-233. doi:10.1016/j.physio.2019.08.014

13. Jarvik JG, Gold LS, Comstock BA, et al. Association of early imaging for back pain with clinical outcomes in older adults. *JAMA*. 2015;313(11):1143-1153. doi:10.1001/jama.2015.1871

14. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. *Ann Rheum Dis*. 1957;16(4):494-502. doi:10.1136/ard.16.4.494

15. Lee SY, Cho NH, Jung YO, Seo YI, Kim HA. Prevalence and risk factors for lumbar spondylosis and its association with low back pain among rural Korean residents. *J Korean Neurosurg Soc*. 2017;60(1):67-74. doi:10.3340/jkns.2016.0505.007

16. Perera RS, Dissanayake PH, Senarath U, Wijayaratne LS, Karunanayake AL, Dissanayake VHW. Associations between disc space narrowing, anterior osteophytes and disability in chronic mechanical low back pain: a cross sectional study. *BMC Musculoskelet Disord*. 2017;18(1):193. doi:10.1186/s12891-017-1562-9

17. Bartlett SJ, Ling SM, Mayo NE, Scott SC, Bingham CO III. Identifying common trajectories of joint space narrowing over two years in knee osteoarthritis. *Arthritis Care Res (Hoboken)*. 2011;63(12):1722-1728. doi:10.1002/acr.20614

18. Felson DT, Gale DR, Elon Gale M, et al. Osteophytes and progression of knee osteoarthritis. *Rheumatology (Oxford)*. 2005;44(1):100-104. doi:10.1093/rheumatology/keh411

19. Vandenbroucke JP, von Elm E, Altman DG, et al; STROBE Initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Ann Intern Med*. 2007;147(8):W163-94. doi:10.7326/0003-4819-147-8-200710160-00010-w1

20. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med*. 2007;147(8):573-577. doi:10.7326/0003-4819-147-8-200710160-00010

21. Hassett G, Hart DJ, Manek NJ, Doyle DV, Spector TD. Risk factors for progression of lumbar spine disc degeneration: the Chingford Study. *Arthritis Rheum*. 2003;48(11):3112-3117. doi:10.1002/art.11321

22. Lane NE, Nevitt MC, Genant HK, Hochberg MC. Reliability of new indices of radiographic osteoarthritis of the hand and hip and lumbar disc degeneration. *J Rheumatol*. 1993;20(11):1911-1918.

23. Co YY, Eaton S, Maxwell MW. The relationship between the St. Thomas and Oswestry Disability Scores and the severity of low back pain. *J Manipulative Physiol Ther*. 1993;16(1):14-18.

24. DAGitty.net. DAGitty—draw and analyze causal diagrams. Accessed April 7, 2021. http://www.dagitty.net/

25. Textor J, van der Zander B, Gilthorpe MS, Liśkiewicz M, Ellison GT. Robust causal inference using directed acyclic graphs: the R package "dagitty". *Int J Epidemiol*. 2016;45(6):1887-1894.

26. Harrell Jr FE. *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. Springer; 2015. doi:10.1007/978-3-319-19425-7

27. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Stürmer T. Variable selection for propensity score models. *Am J Epidemiol*. 2006;163(12):1149-1156. doi:10.1093/aje/kwj149

28. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. Lippincott Williams & Wilkins; 2008.

29. Harrell Jr FE. Package 'Hmisc.' CRAN2018. Accessed July 22, 2020. https://cran.uib.no/web/packages/Hmisc/Hmisc.pdf

30. Al-Rawahi M, Luo J, Pollintine P, Dolan P, Adams MA. Mechanical function of vertebral body osteophytes, as revealed by experiments on cadaveric spines. *Spine (Phila Pa 1976)*. 2011;36(10):770-777. doi:10.1097/BRS.0b013e3181df1a70

31. Brinjikji W, Luetmer PH, Comstock B, et al. Systematic literature review of imaging features of spinal degeneration in asymptomatic populations. *AJNR Am J Neuroradiol*. 2015;36(4):811-816. doi:10.3174/ajnr.A4173

SUPPLEMENT.

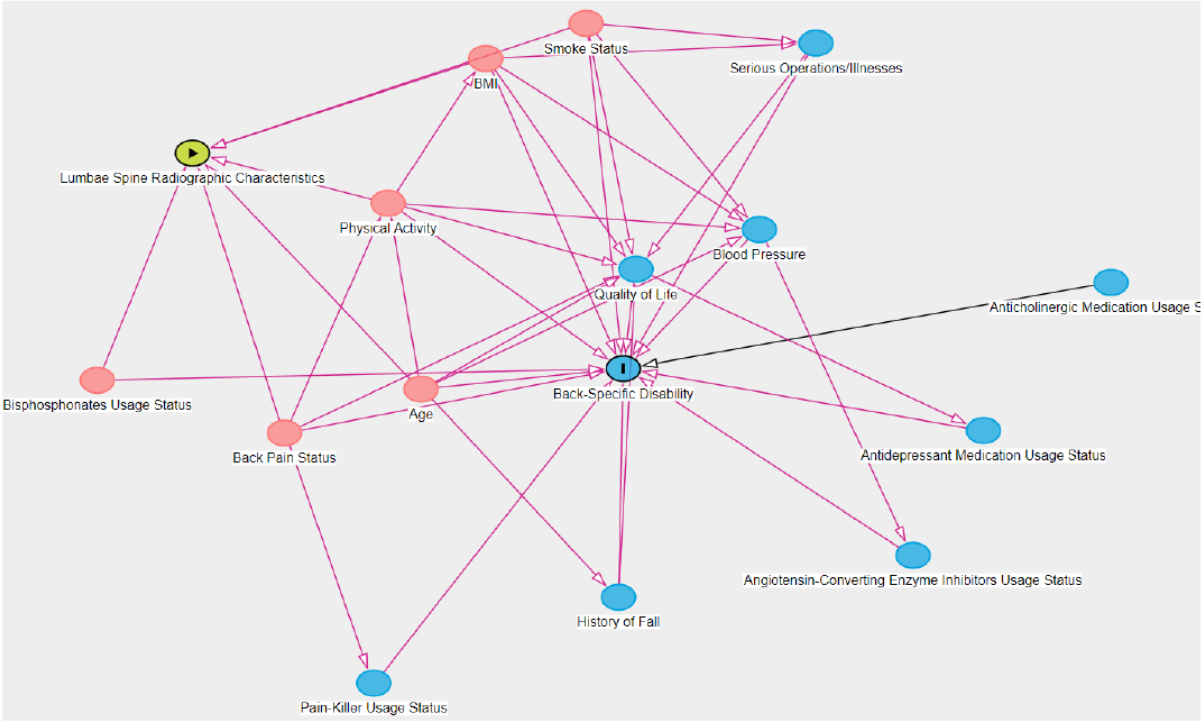
eFigure 1. Causal Diagram
eFigure 2. Histogram of Outcome Distribution
eFigure 3. Missing Data Pattern
eAppendix 1. Methods for Exploratory and Sensitivity Analyses
eAppendix 2. Details of Statistical Methods
eTable 1. Distribution of Lumbar Spine Radiographic Changes at Each Lumbar Spine Segment
eTable 2. Redundancy Analysis of Exposures
eTable 3. Interaction with Age, BMI, or Smoking Status

Supplementary Online Content

Chen L, Perera RS, Radojčić MR, et al. Association of lumbar spine radiographic changes with severity of back pain–related disability among middle-aged, community-dwelling women. *JAMA Netw Open*. 2021;4(5):e2110715. doi:10.1001/jamanetworkopen.2021.10715

This supplementary material has been provided by the authors to give readers additional information about their work.

eFigure 1. Causal Diagram



exposure
outcome
ancestor of exposure
ancestor of outcome
ancestor of exposure *and* outcome

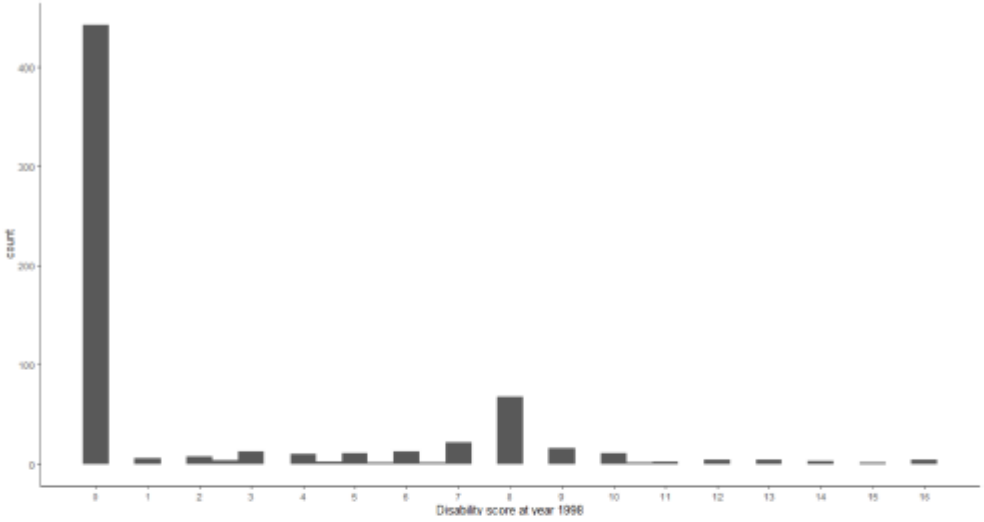
DAGitty is a browser-based environment for creating, editing, and analyzing causal diagrams (also known as directed acyclic graphs or causal Bayesian networks). The focus is on the use of causal diagrams for minimizing bias in empirical studies in epidemiology and other disciplines. We used two steps to choose the final covariates included in the model. At first, we pre-specified a lot of potential confounders based on clinical knowledge. The second step included searching previous literature for well accepted confounders and then drawing the causal diagram to find the minimal sufficient adjustment sets based on the established evidence.

Initially selected covariates: age, BMI, smoke status, back pain status, bisphosphonates usage status, physical activity, blood pressure, serious operations/illnesses, quality of life, history of fall, pain-killer usage status, anticholinergic medication usage status, antidepressant medication usage status, and angiotensin-converting enzyme inhibitors usage status.

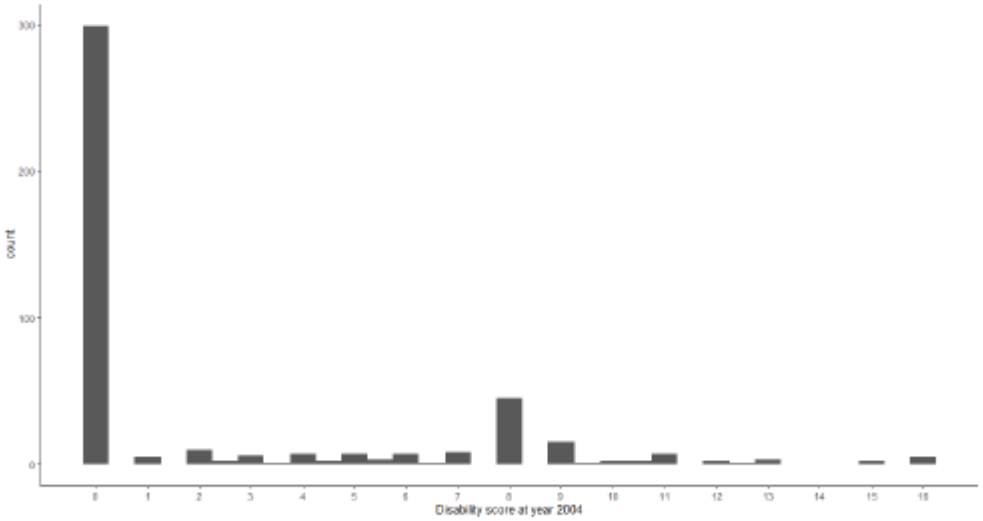
Minimal sufficient adjustment sets: age (as a continuous variable), BMI (as a continuous variable), smoking status (never, current, and ex-smoker), back pain status (yes or no), bisphosphonates use (yes or no), and physical activity (as three domains: walking, job and sport; Walking [per week]: <0.5 miles, 0.5-5 miles, 5-10 miles, and 10+ miles. Job: sedentary, sedentary + occasional exercise, 0.5 sedentary + 0.5 active [or active housework, e.g., daily dust/hoover]), and predominantly manual, active all day. Sport: none, 1 hour per week golf, bowls, badminton, cycling or swimming, 2 hours previous or 1 hour keep-fit, aerobics, squash, and 2 hours + keep-fit, aerobics, squash).

eFigure 2. Histogram of Outcome Distribution

A. Cross-sectional part.

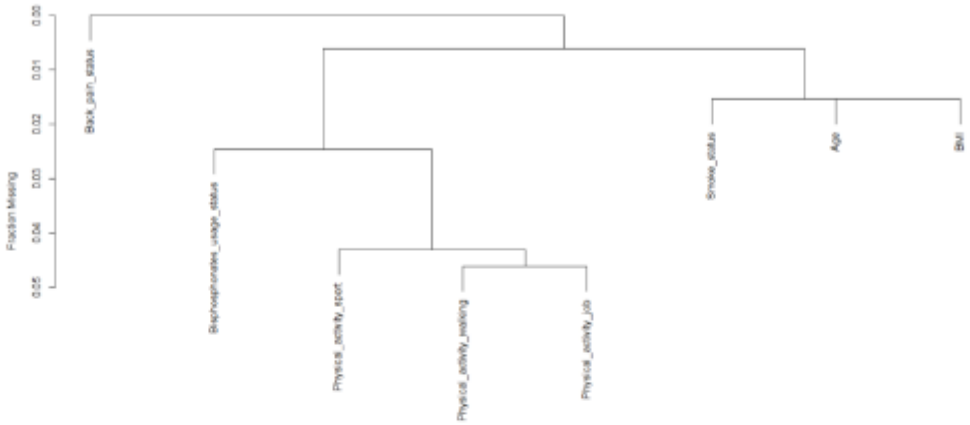


B. Longitudinal part.

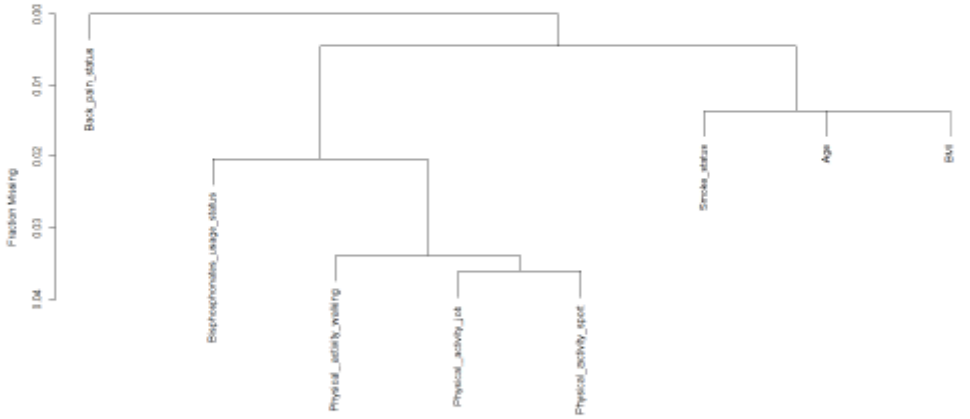


eFigure 3. Missing Data Pattern

A. Cross-sectional part.



B. Longitudinal part.



eAppendix 1. Methods for Exploratory and Sensitivity Analyses

Exploratory analyses

We examined whether the association between lumbar spine radiographic changes and the severity of back pain-related disability differed by age, BMI or smoking status through testing of multiplicative interactions using WALD statistics.

Sensitivity analyses

1) To assess the validity of our cut-off points for defining lumbar spine radiographic changes, we changed the cut-off points from 'no change vs any change' to 'no or mild change vs moderate-to-severe change'.

2) We changed the composite score by using the total original score, which is a sum score from the original grade score at each segment. For example, the K/L grade based on total original score ranged from 0-16 (at each segment 0-4).

3) We built a new composite score based on the disease severity: for disc space narrowing and osteophyte based score, grade 0 vs grade 2 and 3; for K/L grade based score, grade 0 and 1 vs grade 3 and 4.

4) Considering the potential heterogeneity of the population, we restricted our analyses to women with back pain.

5) Considering the potential model misspecification issue, we changed it to linear model and ordinal logistic regression with cloglog link function.

6) To explore the potential influence from unmeasured confounding, E-value was calculated.

7) As one reviewer suggested, we additionally adjusted pain medication and depression considering they might affect the results as potential strong prognostic factors.

eAppendix 2. Details of Statistical Methods

Ordinal logistic regression: Liu et al. indicated that ordinal regression models are robust for continuous outcomes, especially when the distributions of continuous responses are skewed (Stat Med. 2017 Nov 30;36(27):4316-4335). Our continuous outcomes were skewed so that we chose ordinal logistic regression.

Stepped modelling framework: If all covariates were measured at the same time, two models were sufficient: step 1: unadjusted analyses; step 2: analyses adjusted for age, BMI, back pain status, physical activity, bisphosphonates usage status, and smoking status (additionally adjusted for Year 9 back pain-related disability for the longitudinal analysis). The reason for the inclusion of Year 9 back pain-related disability for the longitudinal analysis is that it is a strong prognostic factor for Year 15 back pain-related disability. As VanderWeele et al. indicated, it is often important to control, whenever possible, for the outcome at or prior to the time of the baseline exposure assessment so that confounding control assumption is as plausible as possible (*Statistical Science* 35.3 (2020): 437-466.). In our study, among all selected covariates, physical activity was measured at Year 6 and others were measured at Year 9 (baseline at our study). Considering physical activity could change between Year 6 and Year 9, we built three models: step 1: unadjusted analyses; step 2: analyses adjusted for age, BMI, back pain status, bisphosphonates usage status, and smoking status (additionally adjusted for Year 9 back pain-related disability for the longitudinal analysis); step 3: analyses further adjusted for physical activity.

Exposure modelling: For each type of radiographic changes, our exposure has five values: 0, 1, 2, 3, and 4. The value equals the number of lumbar spine segments affected by radiographic

changes. We could consider the exposure as the unordered categorical variable. In this case, we set 0 as the reference level and obtained the estimate by comparing other values with 0. We could also consider 0-4 as the continuous variable. As a continuous variable, we tested the linear trend by modelling the exposure as continuous variable and reported the P-value (whether the regression coefficient of the exposure variable equalled zero); we also tested the non-linear trend by adding the quartic term to the previous model and reported the P-value (whether the added quartic term could improve performance of the previous model through analysis of variance).

Multiple imputation: It is a general approach to handle missing data in epidemiological and clinical research (BMJ 2009;338:b2393). It includes two steps: step 1: to create multiple copies of the dataset with the missing data replaced by imputed values; step 2: to fit the model to each of the imputed datasets and then calculate the final estimate by combing the estimate from each dataset using Rubin's rule (Rubin, Donald B. *Multiple imputation for nonresponse in surveys*. Vol. 81. John Wiley & Sons, 2004.).

eTable 1. Distribution of Lumbar Spine Radiographic Changes at Each Lumbar Spine Segment

	Cross-section (n=650)	Longitudinal (n=443)
Kellgren-Lawrence grade		
L1-L2		
grade 0	229	160
grade 1	199	140
grade 2	94	57
grade 3	96	68
grade 4	32	18
L2-L3		
grade 0	160	121
grade 1	183	126
grade 2	126	75
grade 3	137	92
grade 4	44	
L3-L4		
grade 0	102	72
grade 1	202	143
grade 2	146	93
grade 3	171	119
grade 4	29	16
L4-L5		
grade 0	135	89
grade 1	230	167
grade 2	88	61
grade 3	116	75
grade 4	81	51
Disc space narrowing		
L1-L2		
grade 0	311	219
grade 1	295	196
grade 2	29	23
grade 3	15	5
L2-L3		
grade 0	300	216
grade 1	296	197
grade 2	40	22
grade 3	14	8
L3-L4		
grade 0	276	198
grade 1	330	222
grade 2	31	15
grade 3	13	8
L4-L5		
grade 0	259	191
grade 1	280	182

grade 2	67	38
grade 3	44	32
Osteophytes		
L1-L2		
grade 0	551	383
grade 1	92	58
grade 2	7	2
grade 3	0	0
L2-L3		
grade 0	509	368
grade 1	129	71
grade 2	10	4
grade 3	2	0
L3-L4		
grade 0	468	335
grade 1	161	98
grade 2	18	10
grade 3	3	0
L4-L5		
grade 0	430	306
grade 1	191	118
grade 2	27	18
grade 3	2	1

eTable 2. Redundancy Analysis of Exposures

R-squared with which each variable can be predicted from all other variables.

R-squared cut-off: 0.75.

	R-squared
Cross-sectional	
L1-L2	
Osteophytes	0.157
Disc space narrowing	0.662
Kellgren-Lawrence grade	0.644
L2-L3	
Osteophytes	0.235
Disc space narrowing	0.649
Kellgren-Lawrence grade	0.633
L3-L4	
Osteophytes	0.198
Disc space narrowing	0.659
Kellgren-Lawrence grade	0.647
L4-L5	
Osteophytes	0.165
Disc space narrowing	0.721
Kellgren-Lawrence grade	0.717
Longitudinal	
L1-L2	
Osteophytes	0.020
Disc space narrowing	0.574
Kellgren-Lawrence grade	0.574
L2-L3	
Osteophytes	0.195
Disc space narrowing	0.616
Kellgren-Lawrence grade	0.603
L3-L4	
Osteophytes	0.147
Disc space narrowing	0.657
Kellgren-Lawrence grade	0.657
L4-L5	
Osteophytes	0.149
Disc space narrowing	0.725
Kellgren-Lawrence grade	0.727

eTable 3. Interaction with Age, BMI, or Smoking Status

	Cross-sectional (P-value)	Longitudinal (P-value)
Kellgren-Lawrence grade based score		
Age	0.84	0.61
BMI	0.62	0.23
Smoke status	0.18	0.66
Osteophytes grade based score		
Age	0.61	0.96
BMI	0.45	0.50
Smoke status	0.63	0.62
Disc space narrowing grade based score		
Age	0.28	0.65
BMI	0.88	0.14
Smoke status	0.41	0.20

eAppendix 3. Sensitivity Analyses Results: Change on the Cut-Off Points of Exposures (corresponding to the first sensitivity analysis; the cut-off points changed from 'no change vs any change' to 'no or mild change vs moderate-to-severe change').

eTable 4. Kellgren-Lawrence Grade–Based Score

Variables	K/L grade based score					P for trend	
	0 segment	1 segment	2 segments	3 segments	4 segments	Linear model	Non-linear model
Cross-sectional (Year 9, n=650)							
Number of women (%)	296 (45.5)	153 (23.5)	99 (15.2)	53 (8.2)	49 (7.5)		
Odds ratio (95% confidence interval)							
Unadjusted	1 (reference)	1.05 (0.70, 1.58)	1.24 (0.78, 1.98)	1.09 (0.60, 1.98)	1.02 (0.53, 1.93)	0.66	0.75
Multivariable adjusted[a]	1 (reference)	1.07 (0.70, 1.64)	1.31 (0.80, 2.13)	1.15 (0.61, 2.16)	1.32 (0.67, 2.62)	0.29	0.78
Further adjusted for physical activity	1 (reference)	1.06 (0.69, 1.63)	1.25 (0.76, 2.04)	1.12 (0.60, 2.12)	1.32 (0.66, 2.64)	0.34	0.89
Longitudinal (Year 15, n=443)							
Number of women (%)	200 (45.1)	110 (24.8)	68 (15.3)	38 (8.6)	27 (6.1)		
Odds ratio (95% confidence interval)							
Unadjusted	1 (reference)	1.16 (0.71, 1.88)	1.14 (0.64, 2.03)	1.34 (0.68, 2.66)	1.39 (0.62, 3.14)	0.29	0.57
Multivariable adjusted[b]	1 (reference)	1.10 (0.64, 1.88)	1.05 (0.54, 2.02)	1.38 (0.65, 2.95)	1.12 (0.45, 2.77)	0.56	0.87
Further adjusted for physical activity	1 (reference)	1.08 (0.62, 1.86)	1.00 (0.51, 1.97)	1.46 (0.66, 3.19)	1.15 (0.46, 2.90)	0.53	0.98

[a] Adjusted for age, BMI, smoke status, back pain status and bisphosphonates usage status.

[b] Adjusted for age, BMI, smoke status, back pain status, bisphosphonates usage status and Year 9 back pain-related disability.

eTable 5. Osteophytes Grade–Based Score

Variables	Osteophytes grade based score					P for trend	
	0 segment	1 segment	2 segments	3 segments	4 segments	Linear model	Non-linear model
Cross-sectional (Year 9, n=650)							
Number of women (%)	596 (91.7)	45 (6.9)	4 (0.6)	4 (0.6)	1 (0.2)		
Odds ratio (95% confidence interval)							
Unadjusted	1 (reference)	0.89 (0.47, 1.69)	0.56 (0.06, 5.02)	N/A	N/A	0.16	0.45
Multivariable adjusted[a]	1 (reference)	0.98 (0.51, 1.89)	0.53 (0.06, 4.89)	N/A	N/A	0.31	0.33
Further adjusted for physical activity	1 (reference)	1.00 (0.52, 1.97)	0.62 (0.07, 5.99)	N/A	N/A	0.35	0.33
Longitudinal (Year 15, n=443)							
Number of women (%)	411 (92.8)	30 (6.8)	1 (0.2)	1 (0.2)	0 (0.0)		
Odds ratio (95% confidence interval)							
Unadjusted	1 (reference)	0.55 (0.23, 1.28)	N/A	N/A	N/A	0.10	0.37
Multivariable adjusted[b]	1 (reference)	0.62 (0.24, 1.56)	N/A	N/A	N/A	0.18	0.85
Further adjusted for physical activity	1 (reference)	0.64 (0.25, 1.62)	N/A	N/A	N/A	0.18	0.88

[a] Adjusted for age, BMI, smoke status, back pain status and bisphosphonates usage status.

[b] Adjusted for age, BMI, smoke status, back pain status, bisphosphonates usage status and Year 9 back pain-related disability.

eTable 6. Disc Space Narrowing Grade–Based Score

Variables	Disc space narrowing grade based score					P for trend	
	0 segment	1 segment	2 segments	3 segments	4 segments	Linear model	Non-linear model
Cross-sectional (Year 9, n=650)							
Number of women (%)	480 (73.8)	114 (17.5)	37 (5.7)	11 (1.7)	8 (1.2)		
Odds ratio (95% confidence interval)							
Unadjusted	1 (reference)	1.39 (0.92, 2.11)	1.01 (0.51, 2.01)	0.75 (0.20, 2.77)	0.33 (0.04, 2.72)	0.87	0.18
Multivariable adjusted[a]	1 (reference)	1.44 (0.93, 2.22)	1.01 (0.49, 2.06)	0.88 (0.23, 3.42)	0.46 (0.05, 3.81)	0.88	0.11
Further adjusted for physical activity	1 (reference)	1.45 (0.93, 2.25)	1.00 (0.48, 2.06)	0.95 (0.24, 3.76)	0.46 (0.05, 3.94)	0.84	0.12
Longitudinal (Year 15, n=443)							
Number of women (%)	332 (74.9)	82 (18.5)	21 (4.7)	5 (1.1)	3 (0.7)		
Odds ratio (95% confidence interval)							
Unadjusted	1 (reference)	1.35 (0.83, 2.22)	1.75 (0.77, 3.97)	0.45 (0.05, 3.88)	N/A	0.64	0.13
Multivariable adjusted[b]	1 (reference)	1.28 (0.73, 2.23)	1.63 (0.67, 3.98)	0.61 (0.06, 6.65)	N/A	0.93	0.06
Further adjusted for physical activity	1 (reference)	1.41 (0.80, 2.49)	1.68 (0.68, 4.12)	0.67 (0.06, 7.59)	N/A	0.72	0.05

[a] Adjusted for age, BMI, smoke status, back pain status and bisphosphonates usage status.

[b] Adjusted for age, BMI, smoke status, back pain status, bisphosphonates usage status and Year 9 back pain-related disability.

eAppendix 4. Sensitivity Analyses Results: Total Original Score of Exposures (corresponding to the second sensitivity analysis; the total original score which is a sum score from the original grade score at each segment).

eTable 7. Kellgren-Lawrence Grade

	Linear model (effect estimate with its 95% confidence interval)	Non-linear model (p-value)
Cross-sectional (Year 9, n=650)		
Unadjusted	-0.12 (-0.34, 0.11)	0.26
Multivariable adjusted[a]	-0.07 (-0.31, 0.17)	0.19
Further adjusted for physical activity	-0.09 (-0.33, 0.16)	0.15
Longitudinal (Year 15, n=443)		
Unadjusted	-0.04 (-0.32, 0.23)	0.11
Multivariable adjusted[b]	-0.10 (-0.41, 0.21)	0.25
Further adjusted for physical activity	-0.11 (-0.43, 0.20)	0.13

[a] Adjusted for age, BMI, smoke status, back pain status and bisphosphonates usage status.

[b] Adjusted for age, BMI, smoke status, back pain status, bisphosphonates usage status and Year 9 back pain-related disability.

eTable 8. Osteophytes Grade

	Linear model	Non-linear model
Cross-sectional (Year 9, n=650)		
Unadjusted	-0.12 (-0.26, 0.01)	1.00
Multivariable adjusted[a]	-0.10 (-0.25, 0.05)	0.70
Further adjusted for physical activity	-0.10 (-0.24, 0.05)	0.71
Longitudinal (Year 15, n=443)		
Unadjusted	-0.25 (-0.44, -0.05)	0.79
Multivariable adjusted[b]	-0.26 (-0.48, -0.05)	0.99
Further adjusted for physical activity	-0.26 (-0.48, -0.04)	0.90

[a] Adjusted for age, BMI, smoke status, back pain status and bisphosphonates usage status.

[b] Adjusted for age, BMI, smoke status, back pain status, bisphosphonates usage status and Year 9 back pain-related disability.

eTable 9. Disc Space Narrowing Grade

	Linear model	Non-linear model
Cross-sectional (Year 9, n=650)		
Unadjusted	-0.07 (-0.30, 0.16)	0.06
Multivariable adjusted[a]	-0.03 (-0.29, 0.22)	0.07
Further adjusted for physical activity	-0.03 (-0.29, 0.22)	0.09
Longitudinal (Year 15, n=443)		
Unadjusted	0.16 (-0.13, 0.44)	0.98
Multivariable adjusted[b]	0.20 (-0.13, 0.52)	0.94
Further adjusted for physical activity	0.25 (-0.08, 0.59)	1.00

[a] Adjusted for age, BMI, smoke status, back pain status and bisphosphonates usage status.

[b] Adjusted for age, BMI, smoke status, back pain status, bisphosphonates usage status and Year 9 back pain-related disability.

eAppendix 5. Sensitivity Analyses Results: Disease Severity (corresponding to the third sensitivity analysis; a new composite score based on the disease severity).

eTable 10. Kellgren-Lawrence Grade–Based Score

Variables	K/L grade based score					P for trend	
	0 segment	1 segment	2 segments	3 segments	4 segments	Linear model	Non-linear model
Cross-sectional (Year 9, n=365)							
Number of women (%)	154 (42.2)	77 (21.1)	47 (12.9)	38 (10.4)	49 (13.4)		
Odds ratio (95% confidence interval)							
Unadjusted	1 (reference)	0.92 (0.52, 1.63)	1.27 (0.66, 2.44)	1.14 (0.56, 2.32)	0.92 (0.47, 1.82)	0.89	0.63
Multivariable adjusted[a]	1 (reference)	1.06 (0.59, 1.92)	1.33 (0.67, 2.65)	1.26 (0.58, 2.70)	1.34 (0.63, 2.84)	0.34	0.83
Further adjusted for physical activity	1 (reference)	1.03 (0.56, 1.88)	1.33 (0.66, 2.66)	1.31 (0.60, 2.89)	1.35 (0.63, 2.88)	0.31	0.87
Longitudinal (Year 15, n=260)							
Number of women (%)	112 (43.1)	60 (23.1)	34 (13.1)	27 (10.4)	27 (10.4)		
Odds ratio (95% confidence interval)							
Unadjusted	1 (reference)	1.31 (0.70, 2.48)	1.49 (0.70, 3.17)	1.15 (0.50, 2.67)	1.25 (0.53, 2.91)	0.51	0.42
Multivariable adjusted[b]	1 (reference)	1.32 (0.65, 2.67)	1.25 (0.52, 2.99)	1.37 (0.54, 3.50)	1.12 (0.42, 2.94)	0.61	0.47
Further adjusted for physical activity	1 (reference)	1.36 (0.66, 2.82)	1.18 (0.47, 2.92)	1.71 (0.63, 4.67)	1.31 (0.47, 3.65)	0.40	0.57

[a] Adjusted for age, BMI, smoke status, back pain status and bisphosphonates usage status.

[b] Adjusted for age, BMI, smoke status, back pain status, bisphosphonates usage status and Year 9 back pain-related disability.

eTable 11. Osteophytes Grade–Based Score

Variables	Osteophytes grade based score					P for trend	
	0 segment	1 segment	2 segments	3 segments	4 segments	Linear model	Non-linear model
Cross-sectional (Year 9, n=283)							
Number of women (%)	258 (91.2)	20 (7.1)	1 (0.4)	3 (1.1)	1 (0.4)		
Odds ratio (95% confidence interval)							
Unadjusted	1 (reference)	0.56 (0.20, 1.56)	2.29 (0.14, 37.06)	N/A	N/A	0.14	0.71
Multivariable adjusted[a]	1 (reference)	0.49 (0.17, 1.39)	1.02 (0.06, 17.96)	N/A	N/A	0.13	0.94
Further adjusted for physical activity	1 (reference)	0.50 (0.17, 1.46)	0.80 (0.04, 16.41)	N/A	N/A	0.11	0.88
Longitudinal (Year 15, n=206)							
Number of women (%)	192 (93.2)	12 (5.8)	1 (0.5)	1 (0.5)	0 (0.0)		
Odds ratio (95% confidence interval)							
Unadjusted	1 (reference)	0.56 (0.15, 2.09)	N/A	N/A	N/A	0.20	0.89
Multivariable adjusted[b]	1 (reference)	0.87 (0.19, 4.00)	N/A	N/A	N/A	0.49	0.88
Further adjusted for physical activity	1 (reference)	1.41 (0.24, 8.33)	N/A	N/A	N/A	0.71	0.87

[a] Adjusted for age, BMI, smoke status, back pain status and bisphosphonates usage status.

[b] Adjusted for age, BMI, smoke status, back pain status, bisphosphonates usage status and Year 9 back pain-related disability.

eTable 12. Disc Space Narrowing Grade–Based Score

Variables	Disc space narrowing grade based score					P for trend	
	0 segment	1 segment	2 segments	3 segments	4 segments	Linear model	Non-linear model
Cross-sectional (Year 9, n=126)							
Number of women (%)	100 (79.4)	11 (8.7)	5 (4.0)	2 (1.6)	8 (6.3)		
Odds ratio (95% confidence interval)							
Unadjusted	1 (reference)	1.69 (0.46, 6.24)	1.45 (0.26, 8.15)	N/A	0.38 (0.04, 3.20)	0.44	0.25
Multivariable adjusted[a]	1 (reference)	1.33 (0.32, 5.50)	2.03 (0.30, 13.90)	N/A	0.53 (0.06, 5.14)	N/A	N/A
Further adjusted for physical activity	1 (reference)	1.20 (0.26, 5.56)	3.25 (0.37, 28.69)	N/A	0.59 (0.05, 6.89)	N/A	N/A
Longitudinal (Year 15, n=87)							
Number of women (%)	70 (80.5)	9 (10.3)	4 (4.6)	1 (1.1)	3 (3.4)		
Odds ratio (95% confidence interval)							
Unadjusted	1 (reference)	1.15 (0.27, 4.99)	0.53 (0.06, 4.98)	1.95 (0.12, 32.49)	N/A	0.39	0.47
Multivariable adjusted[b]	1 (reference)	1.76 (0.17, 18.44)	0.54 (0.03, 9.08)	0.02 (0.0001, 3.28)	N/A	0.33	0.19
Further adjusted for physical activity	1 (reference)	0.69 (0.09, 5.38)	0.06 (0.002, 1.38)	8.07 (0.12, 54.29)	N/A	0.16	0.89

[a] Adjusted for age, BMI, smoke status, back pain status and bisphosphonates usage status.

[b] Adjusted for age, BMI, smoke status, back pain status, bisphosphonates usage status and Year 9 back pain-related disability.

eAppendix 6. Sensitivity Analyses Results: Restricting to Women With Back Pain (corresponding to the fourth sensitivity analysis; the potential heterogeneity of the population).

eTable 13. Kellgren-Lawrence Grade–Based Score

Variables	K/L grade based score					P for trend	
	0 segment	1 segment	2 segments	3 segments	4 segments	Linear model	Non-linear model
Cross-sectional (Year 9, n=210)							
Number of women (%)	51 (24.3)	46 (21.9)	39 (18.6)	42 (20.0)	32 (15.2)		
Odds ratio (95% confidence interval)							
Unadjusted	1 (reference)	0.60 (0.26, 1.38)	0.73 (0.31, 1.71)	0.66 (0.29, 1.54)	0.53 (0.21, 1.33)	0.23	0.74
Multivariable adjusted[a]	1 (reference)	0.54 (0.15, 1.95)	0.29 (0.07, 1.17)	0.77 (0.21, 2.86)	0.21 (0.04, 1.13)	0.14	0.85
Further adjusted for physical activity	1 (reference)	0.74 (0.18, 3.10)	0.23 (0.05, 1.07)	0.76 (0.17, 3.37)	0.24 (0.04, 1.44)	0.14	0.66
Longitudinal (Year 15, n=145)							
Number of women (%)	40 (27.6)	32 (22.1)	24 (16.6)	26 (17.9)	23 (15.9)		
Odds ratio (95% confidence interval)							
Unadjusted	1 (reference)	1.05 (0.40, 2.74)	1.16 (0.42, 3.26)	0.43 (0.14, 1.38)	1.43 (0.52, 3.94)	0.93	0.49
Multivariable adjusted[b]	1 (reference)	0.75 (0.24, 2.36)	1.93 (0.60, 6.15)	0.56 (0.15, 2.12)	1.81 (0.54, 6.10)	0.43	0.62
Further adjusted for physical activity	1 (reference)	0.53 (0.15, 1.85)	1.57 (0.42, 5.78)	0.41 (0.09, 1.75)	1.92 (0.52, 7.06)	0.53	0.32

[a] Adjusted for age, BMI, smoke status and bisphosphonates usage status.

[b] Adjusted for age, BMI, smoke status, back pain status, bisphosphonates usage status and Year 9 back pain-related disability.

eTable 14. Osteophytes Grade–Based Score

Variables	Osteophytes grade based score					P for trend	
	0 segment	1 segment	2 segments	3 segments	4 segments	Linear model	Non-linear model
Cross-sectional (Year 9, n=210)							
Number of women (%)	81 (38.6)	76 (36.2)	32 (15.2)	15 (7.1)	6 (2.9)		
Odds ratio (95% confidence interval)							
Unadjusted	1 (reference)	0.70 (0.36, 1.34)	0.88 (0.38, 2.02)	0.60 (0.18, 2.03)	0.35 (0.04, 3.14)	0.27	0.98
Multivariable adjusted[a]	1 (reference)	1.05 (0.04, 2.76)	1.65 (0.48, 5.62)	0.74 (0.11, 4.83)	N/A	0.79	0.24
Further adjusted for physical activity	1 (reference)	0.97 (0.35, 2.73)	1.49 (0.41, 5.41)	1.08 (0.16, 7.34)	N/A	0.88	0.34
Longitudinal (Year 15, n=145)							
Number of women (%)	56 (38.6)	59 (40.7)	20 (13.8)	8 (5.5)	2 (1.4)		
Odds ratio (95% confidence interval)							
Unadjusted	1 (reference)	0.77 (0.37, 1.63)	0.64 (0.22, 1.87)	0.55 (0.10, 2.88)	N/A	0.18	0.74
Multivariable adjusted[b]	1 (reference)	0.91 (0.39, 2.16)	0.68 (0.20, 2.25)	0.67 (0.10, 4.68)	N/A	0.39	0.66
Further adjusted for physical activity	1 (reference)	0.90 (0.36, 2.25)	0.39 (0.10, 1.51)	0.58 (0.08, 4.29)	N/A	0.18	0.65

[a] Adjusted for age, BMI, smoke status and bisphosphonates usage status.

[b] Adjusted for age, BMI, smoke status, back pain status, bisphosphonates usage status and Year 9 back pain-related disability.

eTable 15. Disc Space Narrowing Grade–Based Score

Variables	Disc space narrowing grade based score					P for trend	
	0 segment	1 segment	2 segments	3 segments	4 segments	Linear model	Non-linear model
Cross-sectional (Year 9, n=210)							
Number of women (%)	25 (11.9)	27 (12.9)	45 (21.4)	41 (19.5)	72 (34.3)		
Odds ratio (95% confidence interval)							
Unadjusted	1 (reference)	0.94 (0.34, 2.64)	0.40 (0.15, 1.10)	0.55 (0.20, 1.47)	0.39 (0.15, 0.99)	0.03	0.58
Multivariable adjusted[a]	1 (reference)	0.81 (0.18, 3.55)	0.29 (0.06, 1.34)	0.55 (0.10, 2.93)	0.32 (0.07, 1.35)	0.10	0.53
Further adjusted for physical activity	1 (reference)	0.43 (0.08, 2.28)	0.23 (0.04, 1.23)	0.39 (0.06, 2.38)	0.26 (0.05, 1.30)	0.19	0.32
Longitudinal (Year 15, n=145)							
Number of women (%)	14 (9.7)	27 (18.6)	31 (21.4)	28 (19.3)	45 (31.0)		
Odds ratio (95% confidence interval)							
Unadjusted	1 (reference)	0.50 (0.15, 1.68)	0.22 (0.06, 0.79)	0.25 (0.07, 0.91)	0.83 (0.28, 2.45)	0.87	0.002
Multivariable adjusted[b]	1 (reference)	0.60 (0.15, 2.39)	0.43 (0.09, 2.00)	0.55 (0.12, 2.45)	1.47 (0.38, 5.75)	0.23	0.04
Further adjusted for physical activity	1 (reference)	0.55 (0.11, 2.66)	0.29 (0.05, 1.61)	0.53 (0.10, 2.80)	1.28 (0.27, 6.07)	0.25	0.03

[a] Adjusted for age, BMI, smoke status and bisphosphonates usage status.

[b] Adjusted for age, BMI, smoke status, back pain status, bisphosphonates usage status and Year 9 back pain-related disability.

eAppendix 7. Sensitivity Analyses Results: Changing the Model to cloglog Link Function (corresponding to the fifth sensitivity analysis; the potential model misspecification).

eTable 16. Kellgren-Lawrence Grade–Based Score

Variables	K/L grade based score					P for trend	
	0 segment	1 segment	2 segments	3 segments	4 segments	Linear model	Non-linear model
Cross-sectional (Year 9, n=650)							
Number of women (%)	154 (23.7)	142 (21.8)	140 (21.5)	118 (18.2)	96 (14.8)		
Odds ratio (95% confidence interval)							
Unadjusted	1 (reference)	1.16 (0.79, 1.70)	0.86 (0.57, 1.30)	0.89 (0.58, 1.37)	0.85 (0.53, 1.34)	0.24	0.82
Multivariable adjusted[a]	1 (reference)	1.20 (0.81, 1.77)	0.86 (0.57, 1.30)	0.93 (0.60, 1.44)	0.91 (0.56, 1.46)	0.36	0.93
Further adjusted for physical activity	1 (reference)	1.18 (0.80, 1.75)	0.84 (0.55, 1.28)	0.89 (0.57, 1.39)	0.89 (0.55, 1.44)	0.30	0.99
Longitudinal (Year 15, n=443)							
Number of women (%)	112 (25.3)	100 (22.6)	97 (21.9)	76 (17.2)	58 (13.1)		
Odds ratio (95% confidence interval)							
Unadjusted	1 (reference)	1.18 (0.75, 1.87)	0.95 (0.59, 1.55)	0.81 (0.48, 1.40)	1.07 (0.62, 1.84)	0.63	0.90
Multivariable adjusted[b]	1 (reference)	0.89 (0.54, 1.46)	0.91 (0.55, 1.53)	0.74 (0.42, 1.31)	0.84 (0.47, 1.52)	0.28	0.60
Further adjusted for physical activity	1 (reference)	0.86 (0.52, 1.43)	0.89 (0.53, 1.49)	0.72 (0.40, 1.28)	0.83 (0.45, 1.50)	0.25	0.54

[a] Adjusted for age, BMI, smoke status, back pain status and bisphosphonates usage status.

[b] Adjusted for age, BMI, smoke status, back pain status, bisphosphonates usage status and Year 9 back pain-related disability.

eTable 17. Osteophytes Grade–Based Score

Variables	Osteophytes grade based score					P for trend	
	0 segment	1 segment	2 segments	3 segments	4 segments	Linear model	Non-linear model
Cross-sectional (Year 9, n=650)							
Number of women (%)	258 (39.7)	226 (34.8)	102 (15.7)	44 (6.8)	20 (3.1)		
Odds ratio (95% confidence interval)							
Unadjusted	1 (reference)	0.84 (0.62, 1.15)	0.84 (0.56, 1.26)	0.59 (0.31, 1.13)	0.83 (0.36, 1.90)	0.12	0.64
Multivariable adjusted[a]	1 (reference)	0.87 (0.63, 1.19)	0.83 (0.55, 1.27)	0.66 (0.34, 1.30)	0.99 (0.42, 2.32)	0.29	0.50
Further adjusted for physical activity	1 (reference)	0.86 (0.63, 1.19)	0.85 (0.56, 1.31)	0.68 (0.35, 1.34)	0.97 (0.41, 2.30)	0.32	0.55
Longitudinal (Year 15, n=443)							
Number of women (%)	192 (43.3)	157 (35.4)	67 (15.1)	19 (4.3)	8 (1.8)		
Odds ratio (95% confidence interval)							
Unadjusted	1 (reference)	0.77 (0.53, 1.11)	0.68 (0.41, 1.13)	0.49 (0.18, 1.34)	0.58 (0.14, 2.37)	0.04	0.71
Multivariable adjusted[b]	1 (reference)	0.91 (0.62, 1.33)	0.68 (0.40, 1.16)	0.52 (0.18, 1.45)	0.48 (0.11, 2.06)	0.03	0.97
Further adjusted for physical activity	1 (reference)	0.94 (0.63, 1.39)	0.63 (0.36, 1.10)	0.55 (0.19, 1.57)	0.54 (0.12, 2.33)	0.03	0.93

[a] Adjusted for age, BMI, smoke status, back pain status and bisphosphonates usage status.

[b] Adjusted for age, BMI, smoke status, back pain status, bisphosphonates usage status and Year 9 back pain-related disability.

eTable 18. Disc Space Narrowing Grade–Based Score

Variables	Disc space narrowing grade based score					P for trend	
	0 segment	1 segment	2 segments	3 segments	4 segments	Linear model	Non-linear model
Cross-sectional (Year 9, n=650)							
Number of women (%)	100 (15.4)	107 (16.5)	147 (22.6)	131 (20.2)	165 (25.4)		
Odds ratio (95% confidence interval)							
Unadjusted	1 (reference)	1.28 (0.79, 2.10)	1.26 (0.79, 1.99)	1.31 (0.82, 2.10)	0.99 (0.62, 1.58)	0.80	0.10
Multivariable adjusted[a]	1 (reference)	1.36 (0.83, 2.25)	1.49 (0.92, 2.39)	1.42 (0.88, 2.30)	1.10 (0.67, 1.81)	0.87	0.04
Further adjusted for physical activity	1 (reference)	1.33 (0.80, 2.21)	1.47 (0.91, 2.38)	1.40 (0.86, 2.27)	1.06 (0.65, 1.75)	0.98	0.03
Longitudinal (Year 15, n=443)							
Number of women (%)	70 (15.8)	84 (19.0)	102 (23.0)	88 (19.9)	99 (22.3)		
Odds ratio (95% confidence interval)							
Unadjusted	1 (reference)	0.90 (0.52, 1.57)	0.79 (0.46, 1.36)	1.01 (0.59, 1.73)	1.15 (0.69, 1.93)	0.42	0.21
Multivariable adjusted[b]	1 (reference)	0.86 (0.48, 1.54)	0.89 (0.50, 1.59)	1.08 (0.61, 1.90)	1.37 (0.78, 2.39)	0.13	0.23
Further adjusted for physical activity	1 (reference)	0.86 (0.46, 1.59)	0.84 (0.46, 1.54)	1.16 (0.64, 2.10)	1.51 (0.84, 2.73)	0.08	0.14

[a] Adjusted for age, BMI, smoke status, back pain status and bisphosphonates usage status.

[b] Adjusted for age, BMI, smoke status, back pain status, bisphosphonates usage status and Year 9 back pain-related disability.

eAppendix 8. Sensitivity Analyses Results: Change the Model to Linear Regression (corresponding to the fifth sensitivity analysis; the potential model misspecification).

eTable 19. Kellgren-Lawrence Grade–Based Score

Variables	K/L grade based score					P for trend	
	0 segment	1 segment	2 segments	3 segments	4 segments	Linear model	Non-linear model
Cross-sectional (Year 9, n=650)							
Number of women (%)	154 (23.7)	142 (21.8)	140 (21.5)	118 (18.2)	96 (14.8)		
Odds ratio (95% confidence interval)							
Unadjusted	1 (reference)	1.42 (0.59, 3.40)	0.86 (0.36, 2.06)	0.99 (0.40, 2.50)	0.65 (0.25, 1.74)	0.31	0.49
Multivariable adjusted[a]	1 (reference)	1.44 (0.61, 3.42)	0.88 (0.37, 2.09)	1.01 (0.40, 2.55)	0.73 (0.27, 1.99)	0.43	0.55
Further adjusted for physical activity	1 (reference)	1.40 (0.59, 3.36)	0.81 (0.34, 1.94)	0.96 (0.38, 2.42)	0.69 (0.25, 1.88)	0.35	0.61
Longitudinal (Year 15, n=443)							
Number of women (%)	112 (25.3)	100 (22.6)	97 (21.9)	76 (17.2)	58 (13.1)		
Odds ratio (95% confidence interval)							
Unadjusted	1 (reference)	1.41 (0.48, 4.09)	0.91 (0.31, 2.68)	0.56 (0.18, 1.77)	0.89 (0.25, 3.13)	0.38	0.93
Multivariable adjusted[b]	1 (reference)	0.93 (0.36, 2.44)	1.01 (0.38, 2.67)	0.60 (0.21, 1.74)	0.54 (0.17, 1.75)	0.14	0.89
Further adjusted for physical activity	1 (reference)	0.92 (0.35, 2.43)	1.00 (0.38, 2.68)	0.59 (0.20, 1.72)	0.54 (0.16, 1.74)	0.13	0.87

[a] Adjusted for age, BMI, smoke status, back pain status and bisphosphonates usage status.

[b] Adjusted for age, BMI, smoke status, back pain status, bisphosphonates usage status and Year 9 back pain-related disability.

eTable 20. Osteophytes Grade–Based Score

Variables	Osteophytes grade based score					P for trend	
	0 segment	1 segment	2 segments	3 segments	4 segments	Linear model	Non-linear model
Cross-sectional (Year 9, n=650)							
Number of women (%)	258 (39.7)	226 (34.8)	102 (15.7)	44 (6.8)	20 (3.1)		
Odds ratio (95% confidence interval)							
Unadjusted	1 (reference)	0.64 (0.32, 1.27)	0.64 (0.26, 1.54)	0.35 (0.26, 1.20)	0.68 (0.12, 3.90)	0.11	0.47
Multivariable adjusted[a]	1 (reference)	0.68 (0.35, 1.34)	0.61 (0.25, 1.48)	0.42 (0.12, 1.46)	0.92 (0.16, 5.43)	0.22	0.36
Further adjusted for physical activity	1 (reference)	0.67 (0.34, 1.32)	0.63 (0.26, 1.55)	0.43 (0.12, 1.52)	0.84 (0.14, 5.01)	0.21	0.41
Longitudinal (Year 15, n=443)							
Number of women (%)	192 (43.3)	157 (35.4)	67 (15.1)	19 (4.3)	8 (1.8)		
Odds ratio (95% confidence interval)							
Unadjusted	1 (reference)	0.46 (0.20, 1.05)	0.28 (0.09, 0.83)	0.20 (0.03, 1.25)	0.18 (0.01, 2.96)	0.004	0.47
Multivariable adjusted[b]	1 (reference)	0.58 (0.28, 1.24)	0.27 (0.10, 0.74)	0.32 (0.06, 1.74)	0.12 (0.01, 1.41)	0.002	0.61
Further adjusted for physical activity	1 (reference)	0.59 (0.27, 1.26)	0.27 (0.10, 0.76)	0.32 (0.06, 1.76)	0.13 (0.01, 1.63)	0.002	0.57

[a] Adjusted for age, BMI, smoke status, back pain status and bisphosphonates usage status.

[b] Adjusted for age, BMI, smoke status, back pain status, bisphosphonates usage status and Year 9 back pain-related disability.

eTable 21. Disc Space Narrowing Grade–Based Score

Variables	Disc space narrowing grade based score					P for trend	
	0 segment	1 segment	2 segments	3 segments	4 segments	Linear model	Non-linear model
Cross-sectional (Year 9, n=650)							
Number of women (%)	100 (15.4)	107 (16.5)	147 (22.6)	131 (20.2)	165 (25.4)		
Odds ratio (95% confidence interval)							
Unadjusted	1 (reference)	1.59 (0.56, 4.52)	1.68 (0.63, 4.45)	1.43 (0.53, 3.90)	0.92 (0.35, 2.39)	0.62	0.11
Multivariable adjusted[a]	1 (reference)	1.70 (0.60, 4.82)	2.07 (0.78, 5.46)	1.52 (0.56, 4.12)	0.98 (0.36, 2.63)	0.69	0.05
Further adjusted for physical activity	1 (reference)	1.61 (0.56, 4.61)	2.00 (0.75, 5.34)	1.50 (0.55, 4.10)	0.94 (0.35, 2.56)	0.65	0.05
Longitudinal (Year 15, n=443)							
Number of women (%)	70 (15.8)	84 (19.0)	102 (23.0)	88 (19.9)	99 (22.3)		
Odds ratio (95% confidence interval)							
Unadjusted	1 (reference)	0.96 (0.27, 3.36)	0.75 (0.23, 2.51)	1.55 (0.45, 5.36)	1.83 (0.55, 6.15)	0.18	0.36
Multivariable adjusted[b]	1 (reference)	0.64 (0.20, 2.01)	0.67 (0.22, 2.03)	1.22 (0.39, 3.77)	1.35 (0.43, 4.31)	0.14	0.22
Further adjusted for physical activity	1 (reference)	0.60 (0.19, 1.95)	0.63 (0.20, 1.92)	1.19 (0.38, 3.72)	1.39 (0.43, 4.49)	0.14	0.18

[a] Adjusted for age, BMI, smoke status, back pain status and bisphosphonates usage status.

[b] Adjusted for age, BMI, smoke status, back pain status, bisphosphonates usage status and Year 9 back pain-related disability.

eTable 22. Additionally Adjusted for Pain Medication and Depression (corresponding to the seventh sensitivity analysis; potential strong prognostic factors).

	Cross-sectional (n=650)					
	0 segment (n=154)	1 segment (n=142)	2 segments (n=140)	3 segments (n=118)	4 segments (n=96)	Whole (n=650)
Pain Medication						
Yes	10 (6.5)	7 (4.9)	16 (11.4)	8 (6.8)	10 (10.4)	51 (7.8)
No	68 (44.2)	60 (42.3)	59 (42.1)	53 (44.9)	37 (38.5)	277 (42.6)
Missing	76 (49.4)	75 (52.8)	65 (46.4)	57 (48.3)	49 (51.1)	322 (49.5)
Depression						
Yes	0 (0.0)	1 (0.7)	0 (0.0)	1 (0.8)	1 (1.0)	3 (0.5)
No	154 (100.0)	141 (99.3)	140 (100.0)	117 (99.2)	95 (99.0)	647 (99.5)

	Longitudinal (n=443)					
	0 segment (n=112)	1 segment (n=100)	2 segments (n=97)	3 segments (n=76)	4 segments (n=58)	Whole (n=443)
Pain Medication						
Yes	10 (8.9)	6 (6.0)	9 (9.3)	4 (5.3)	6 (10.3)	35 (7.9)
No	42 (37.5)	42 (42.0)	48 (49.5)	35 (46.1)	24 (41.4)	191 (43.1)

Missing	60 (53.6)	52 (52.0)	40 (41.2)	37 (48.7)	28 (48.3)	217 (49.0)
Depression						
Yes	0 (0.0)	1 (1.0)	1 (1.0)	1 (1.3)	1 (1.7)	3 (0.7)
No	112 (100.0)	99 (99.0)	96 (99.0)	75 (98.7)	57 (98.3)	443 (99.3)

The columns except the first correspond to the number of segments of Lumbar Spine Radiographic Changes (Kellgren-Lawrence grade based). Data are present as number (percentage) of participants unless otherwise indicated.

Women reported current medication use in an open field question within the medical history questionnaire. Data on use of non-opioid and opioid analgesics, defined based on Anatomical Therapeutic Chemical codes M01 and N02, were extracted from this question from Year 9. The details are:

With specific name

With opioid involved

Dihydrocodeine; Dextromoramide; Tramadol; Codeine; Morphine; Paracetamol and dextropropoxyphene; Paracetamol and Codeine; Paracetamol and dihydrocodeine

Without opioid involved

Indomethacin; Ibuprofen; Diclofenac; Etodolac; Fenbufen; Flurbiprofen; Fenoprofen; Mefenamic acid; Naproxen; Piroxicam; Ketoprofen; Movelat; Glucosamine; Feverfew;

Paracetamol

Without specific name

NSAID; Anti-inflammatory; Analgesics; Painkillers

Depression was defined by text response. From Year 1 to Year 4, women were asked the question: Serious operations/illnesses: Other? From Year 8 to Year 9, women were asked the question: Any major illnesses or operations? If the participant reported depression in at least one year (Year 1 to 9, our baseline is Year 9), we defined the value of this covariate as yes.

eTable 23. Kellgren-Lawrence Grade–Based Score

Variables	K/L grade based score					P for trend	
	0 segment	1 segment	2 segments	3 segments	4 segments	Linear model	Non-linear model
Cross-sectional (Year 9, n=650)							
Number of women (%)	154 (23.7)	142 (21.8)	140 (21.5)	118 (18.2)	96 (14.8)		
Odds ratio (95% confidence interval)							
Additional adjusted[a]	1 (reference)	1.21 (0.61, 2.42)	0.85 (0.42, 1.75)	1.25 (0.60, 2.63)	0.87 (0.38, 2.02)	0.84	0.76
Longitudinal (Year 15, n=443)							
Number of women (%)	112 (25.3)	100 (22.6)	97 (21.9)	76 (17.2)	58 (13.1)		
Odds ratio (95% confidence interval)							
Additional adjusted[b]	1 (reference)	1.10 (0.59, 2.04)	0.92 (0.49, 1.74)	0.66 (0.33, 1.33)	0.78 (0.37, 1.62)	0.14	0.86

[a] Adjusted for age, BMI, smoke status, back pain status, bisphosphonates usage status, physical activity, pain medication and depression.

[b] Adjusted for age, BMI, smoke status, back pain status, bisphosphonates usage status, physical activity, pain medication, depression and Year 9 back pain-related disability.

eTable 24. Osteophytes Grade–Based Score

Variables	Osteophyte grade based score					P for trend	
	0 segment	1 segment	2 segments	3 segments	4 segments	Linear model	Non-linear model
Cross-sectional (Year 9, n=650)							
Number of women (%)	154 (23.7)	142 (21.8)	140 (21.5)	118 (18.2)	96 (14.8)		
Odds ratio (95% confidence interval)							
Additional adjusted[a]	1 (reference)	1.17 (0.67, 2.05)	1.25 (0.63, 2.48)	0.66 (0.22, 1.99)	2.46 (0.70, 8.56)	0.52	0.86
Longitudinal (Year 15, n=443)							
Number of women (%)	112 (25.3)	100 (22.6)	97 (21.9)	76 (17.2)	58 (13.1)		
Odds ratio (95% confidence interval)							
Additional adjusted[b]	1 (reference)	0.87 (0.44, 1.73)	0.54 (0.22, 1.32)	0.29 (0.05, 1.63)	0.38 (0.06, 2.40)	0.04	0.87

[a] Adjusted for age, BMI, smoke status, back pain status, bisphosphonates usage status, physical activity, pain medication and depression.

[b] Adjusted for age, BMI, smoke status, back pain status, bisphosphonates usage status, physical activity, pain medication, depression and Year 9 back pain-related disability.

eTable 25. Disc Space Narrowing Grade–Based Score

Variables	Disc space narrowing grade based score					P for trend	
	0 segment	1 segment	2 segments	3 segments	4 segments	Linear model	Non-linear model
Cross-sectional (Year 9, n=650)							
Number of women (%)	154 (23.7)	142 (21.8)	140 (21.5)	118 (18.2)	96 (14.8)		
Odds ratio (95% confidence interval)							
Additional adjusted[a]	1 (reference)	1.42 (0.58, 3.45)	1.40 (0.59, 3.32)	1.48 (0.62, 3.54)	1.20 (0.50, 2.86)	0.86	0.34
Longitudinal (Year 15, n=443)							
Number of women (%)	112 (25.3)	100 (22.6)	97 (21.9)	76 (17.2)	58 (13.1)		
Odds ratio (95% confidence interval)							
Additional adjusted[b]	1 (reference)	0.62 (0.21, 1.83)	0.45 (0.15, 1.35)	0.87 (0.29, 2.55)	1.07 (0.38, 3.02)	0.49	0.10

[a] Adjusted for age, BMI, smoke status, back pain status, bisphosphonates usage status, physical activity, pain medication and depression.

[b] Adjusted for age, BMI, smoke status, back pain status, bisphosphonates usage status, physical activity, pain medication, depression and Year 9 back pain-related disability.

eTable 26. E-value (corresponding to the sixth sensitivity analysis; potential influence from unmeasured confounding).

Variable	Compared with 0 segment			
	1 segment	2 segments	3 segments	4 segments
Cross-sectional				
Kellgren-Lawrence grade based score				
Odds ratio (95%CI)	1.22 (0.76, 1.96)	0.84 (0.51, 1.38)	0.92 (0.54,1.56)	0.89 (0.50, 1.57)
E-value (lower Bound)	1.44 (1)	1.41 (1)	1.25 (1)	1.31 (1)
Osteophytes grade based score				
Odds ratio (95%CI)	0.83 (0.57, 1.22)	0.78 (0.47, 1.30)	0.58 (0.27, 1.26)	1.03 (0.37, 2.85)
E-value (lower Bound)	1.43 (1)	1.52 (1)	1.95 (1)	1.14 (1)
Disc space narrowing grade based score				
Odds ratio (95%CI)	1.43 (0.78, 2.61)	1.56 (0.88, 2.76)	1.44 (0.81, 2.57)	1.07 (0.60, 1.92)
E-value (lower Bound)	1.68 (1)	1.81 (1)	1.69 (1)	1.22 (1)
Longitudinal				
Kellgren-Lawrence grade based score				
Odds ratio (95%CI)	1.06 (0.57, 1.96)	0.94 (0.50, 1.76)	0.69 (0.34, 1.38)	0.83 (0.40, 1.72)
E-value (lower Bound)	1.20 (1)	1.21 (1)	1.70 (1)	1.43 (1)
Osteophytes grade based score				
Odds ratio (95%CI)	0.76 (0.47, 1.24)	0.53 (0.28, 1.02)	0.49 (0.14, 1.70)	0.31 (0.06, 1.72)
E-value (lower Bound)	1.56 (1)	2.09 (1.21)	2.21 (1)	2.99 (1)
Disc space narrowing grade based score				
Odds ratio (95%CI)	0.72 (0.34, 1.53)	0.74 (0.36, 1.52)	1.06 (0.52, 2.20)	1.26 (0.62, 2.57)
E-value (lower Bound)	1.64 (1)	1.60 (1)	1.20 (1)	1.42 (1)

We calculated E-value through Online Calculator (https://mmathur.shinyapps.io/evalue/) based on results from step 2 of the stepped modelling framework.

Explanation: for an unmeasured confounder to explain the OR estimate of 1.22, the unmeasured confounder would have to be associated with both the exposure and the outcome by 1.44-fold above and beyond the measured confounders.

CHAPTER SIX

Association of analgesic use trajectories with mortality and quality of life in middle-aged, community-dwelling women: a population-based cohort study.

Chapter Six has been submitted for publication as:

Chen L., Radojčić M.R., Perera R.S., Beckenkamp P.R., Ferreira P.H., Hart D.J., Spector T.D., Arden N.K., Ferreira M.L. Association of analgesic use trajectories with mortality and quality of life in middle-aged, community-dwelling women: a population-based cohort study. *Under review*

Statement from co-authors confirming authorship contribution of the PhD candidate

The co-authors of the paper "Association of analgesic use trajectories with mortality and quality of life in middle-aged, community-dwelling women: a population-based cohort study" confirm that Lingxiao Chen has made the following contributions:

Conception and design of the research
Analysis and interpretation of the findings
Writing of the manuscript and critical appraisal of the content

In addition to the statements above, in cases where I am not the corresponding author of a published item, permission to include the published material has been granted by the corresponding author.

Lingxiao Chen Date: 21 July 2021

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

Manuela Loureiro Ferreira Date: 21 July 2021

Association of analgesic use trajectories with mortality and quality of life in middle-aged, community-dwelling women: a population-based cohort study

Lingxiao Chen, MMed[1]; Maja R Radojčić, PhD[2,3]; Romain S Perera, PhD[2,4]; Paula R Beckenkamp, PhD[5]; Paulo H Ferreira, PhD[5]; Deborah J Hart, PhD[6]; Tim D Spector, PhD[6]; Nigel K Arden, PhD[2,7*]; Manuela L Ferreira, PhD[1*]

1. Institute of Bone and Joint Research, The Kolling Institute, Northern Clinical School, Faculty of Medicine and Health, University of Sydney, Sydney, Australia

2. Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, United Kingdom

3. Centre for Sport, Exercise and Osteoarthritis Research Versus Arthritis, University of Oxford, Oxford, United Kingdom

4. Department of Allied Health Sciences, Faculty of Medicine, University of Colombo, Colombo, Sri Lanka

5. University of Sydney, School of Health Sciences, Faculty of Medicine and Health, Sydney, New South Wales, Australia

6. Department of Twin Research and Genetic Epidemiology, King's College London, London, United Kingdom

7. MRC Environmental Epidemiology Unit, University of Southampton, Southampton, United Kingdom

*These authors contributed equally as senior authors.

Corresponding Author: Lingxiao Chen, MBBS, MMed, Institute of Bone and Joint Research, The Kolling Institute, Northern Clinical School, Faculty of Medicine and Health, University of

Sydney, Sydney, NSW 2064, Australia (lche4036@uni.sydney.edu.au; @LingxiaoChen2 on Twitter; Tel: 610466965326)

ABSTRACT

IMPORTANCE Analgesics are widely used and at a steadily increasing rate over the past years. Given the many associated adverse events of analgesics and potential impact their long-term use may have on the patient's general health, it is relevant to better understand whether different patterns of analgesic use may influence adverse outcomes.

OBJECTIVE To identify distinct trajectories of analgesic use and identify the association of these trajectories with mortality and quality of life.

DESIGN A population-based prospective cohort.

SETTING The Chingford 1000 Women Study, UK.

PARTICIPANTS Middle-aged, community-dwelling women. Data were collected from Year 1 (1989 to 1991) to 15th August 2014.

EXPOSURES Reported use of analgesics (anatomical therapeutic chemical codes: M01 and N02) as presented in the participant's medical history questionnaire.

MAIN OUTCOMES AND MESSURES All-cause mortality, cause-specific mortality, and quality of life.

RESULTS Among 804 women (mean [SD] age, 62.7 [5.9] years; Year 10 [1998 to 2000]), we identified three distinct trajectories of analgesic use: (i) 'no use' group (691, 85.9%); (ii) 'increasing probability to use' group (73, 9.1%); and (iii) 'constant analgesic use' group (40, 5.0%). Compared with the 'no use' group, the 'constant analgesic use' group was associated with 2.15 times higher risk of all-cause mortality (95% confidence interval [CI]: 1.18 to 3.91) using a multivariable Cox proportional hazard model controlling for selected covariates. No association between cause-specific mortality and pattern of analgesic use was found. Worse quality of life in terms of physical function, role limitations due to physical health and pain was associated with constant and high probability and increased probability of using analgesics.

CONCLUSIONS AND RELEVANCE In this cohort of middle-aged women, a small group of women had a high and constant probability of using analgesic over the study period and a markedly higher risk of all-cause mortality compared to those with no or low probability of using analgesics.

Introduction

Chronic pain is common. A meta-analysis which identified 122 publications in 28 low-income and middle-income countries indicated that the prevalence of chronic pain was 33% in the general adult population and 56% in the general elderly population[1]. Based on the estimate from the Centers for Disease Control and Prevention, about one fifth (50.0 million) of the US adult population reported chronic pain and 8% (19.6 million) reported chronic pain that frequently limits life or work activities in 2016[2]. Pharmacologic treatments including opioid, nonsteroidal anti-inflammatory drugs (NSAIDs) and acetaminophen are mainstream options to alleviate pain, although the effects of these treatments on pain and function are small to moderate[3-5].

Despite the many risks associated with the use of analgesics, the duration of use and type of analgesic used may vary substantially in people with chronic pain[3,6]. A study conducted in Denmark found that only a minor percentage (5.3%) of patients who used opioid preoperatively still used it with same dose after receiving total knee arthroplasty[7]. Another Norwegian study showed that about two thirds (65%) of patients who started to use non-opioid analgesics escalated to weak or strong opioids within the five-year follow-up[8]. These differences might provide distinct prognostic information considering the mortality and quality of life[9]. Previous studies attempting to elucidate the trajectory of analgesic use have two main limitations: firstly, few studies have reported on long-term outcomes, especially on mortality and quality of life[10,11]; and, secondly, past studies have failed to account for the potential induction period[12]. This is because the exposure status at a given time will correlate with a possible increase or decrease in disease only at some later time, which might introduce bias if we modelled the exposure-outcome association without considering the later time issue; lag period analysis could assess the potential influence by the induction period[12].

The aim of our study was: 1) to identify distinct trajectories of analgesic use in middle-aged, community-dwelling women; and 2) to identify whether these trajectories are associated with increased risk of mortality and worse quality of life.

Methods

Study Sample

From an age/sex register of a large practice of over 11000 patients in Chingford, outer London (UK), all 1353 women in the age range 45-64 years were invited to participate in a study assessing musculoskeletal disease in the population. A total of 1003 women were examined between 1989 and 1991 (Year 1, baseline visit); six died, 66 had moved away and 278 refused or did not respond. All the women lived within five miles of the general practice, and 98% of the women were white. Women from this general practice are similar to the UK general population in terms of weigh, height and BMI[13]. Only participants who reported at least three out of seven waves of data about analgesic use were included (Appendix S1). Based on analysis framework from previous studies[14,15], we excluded women who died before Year 10 and those who did not attend Year 10 visit. The Waltham Forest and Redbridge local research ethics committee has approved the study, and all participants provided written informed consent to participate in the study.

Analgesic Assessment

Women reported current medication use in an open field question within the medical history questionnaire. Data on use of non-opioid and opioid analgesics, defined based on Anatomical Therapeutic Chemical codes M01 and N02, were extracted from this question from follow-up years 1, 2, 3, 4, 8, 9 and 10 (Appendix S2). Penicillamine were excluded as these are not primarily prescribed as pain medication. Likewise, Aspirin was excluded as the dose data were insufficient to determine the purpose of their use (i.e., for pain relief or control of existing cardiovascular disease).

Mortality and Quality of Life

For all-cause mortality, which was the primary outcome, study participants were followed from the clinical date at Year 10 (1998 to 2000) visit and continued until death, loss to follow-up, or the end of the follow-up on 15th August 2014. The Health and Social Care Information Centre provided detailed mortality information based on the information collected by the Office for National Statistics from civil registration records. Cause-specific mortality, based on information from death certificates, was divided into cancer-related, cardiovascular disease-related, and others.

Quality of life was measured at Year 15 using the 36-Item Short-Form Health Survey (SF-36 short) which is widely used for routine monitoring and assessment of care outcomes in adult [16]. The SF-36 comprises of eight subscales: physical function, bodily pain, role limitations due to physical health problems, role limitations due to personal or emotional problems, general mental health, social function, energy/fatigue, and general health perceptions[17]. The eight subscales were scored following the instructions for RAND 36-Item Health Survey 1.0: scores range from 0 to 100, with higher scores indicating better quality of life. The minimal clinically important difference (MCID) was defined at 10 points for the lower confidence limit of a positive value and the upper confidence limit of a negative value[18,19].

Covariates

Based on Global Burden of Diseases, Injuries, and Risk Factors Study 2019 (systematically assessed 87 risk factors for mortality in 204 countries and territories) and a previous Chingford study which also used mortality as the outcome, the following covariates were selected[20,21].

1) measurement at Year 10 visit: age (continuous), systolic blood pressure (continuous), body mass index (continuous), smoking status (never, current, and ex-smoker), fasting blood glucose (continuous), cholesterol level (total cholesterol, continuous) and major illness or operations (e.g., cancer, cardiovascular disease, gastrectomy, and cholecystectomy; summarized as a binary variable, yes or no; the definition details in Appendix S3).

2) measurement before Year 10 visit: frequency of alcohol consumption (never, weekly, and social occasions), physical activity participation (from the question "were you a physically active person at age 30?"; yes or no), and occupation (manual or non-manual).

Statistical Analysis

We used a group-based trajectory model to determine trajectory groups of analgesic use[22]. The TRAJ package in Stata (version 15.1) was used to fit logistic model with up to cubic function and test trajectory groups[23]. Bayesian information criterion and posterior probability (>0.70) criteria were also used to determine the optimal number and shapes of trajectory groups, as previously

described[24]. The model with three trajectories and a quadratic function of follow-up year showed the best fit to the data. The average of the posterior probabilities of group membership for individuals assigned to each group was 0.84, 0.93, and 0.90, which indicated a good adequacy of the selected model[25]. Descriptive statistics were performed for each covariate in each trajectory group.

To identify whether these trajectories are associated with increased risk of mortality and quality of life, trajectory groups were treated as categorical exposure variables with 'no use' group as a reference. For all-cause mortality, hazard ratios (HR) and 95% confidence intervals (CI) were estimated through Cox proportional hazards model. Proportional hazard assumption was met. Cause-specific mortality was calculated using multistate survival analysis considering competing risks (e.g., women who died due to cancer could not die due to cardiovascular disease)[26]. Quality of life was calculated for each subscale through linear regression and reported as mean differences (MD) and 95% CI. Multiple imputation was used to deal with missing data in covariates with missing at random assumption[27]. Different measurement times for the covariates was adjusted using the following three models for each analysis: model 1: adjusted for age; model 2: adjusted for age, systolic blood pressure, body mass index, smoking status, glucose, cholesterol level and major illness or operations; model 3: further adjusted for alcohol consumption, physical activity participation, and occupation. Effect estimates from model 2 were presented in the results section.

For the primary outcome (all-cause mortality), e-value was also calculated to assess the influence from unmeasured confoundings[28]. To assess the influence of potential induction period, we used different lag time periods (3, 5, 7, and 10-year lag)[12]. To explore the effect estimate of non-opioid vs opioid analgesics, we performed four sensitivity analyses (Appendix S4). Considering severe cancers could cause extreme pain which might bias the results, we excluded women with cancer as a sensitivity analysis (Appendix S4). All statistical analyses, except identifying trajectory groups, were performed in R (R Core Team, version 4.0.2).

Results

A total of 1003 women were included in the Chingford 1000 Women study. We excluded 48 women who reported one or two waves data for analgesic use and 151 women who did not attend the Year 10 visit or died before the Year 10 visit. These exclusions resulted in a final sample of 804 women. Based on group-based trajectory modelling, we identified three distinct trajectories of analgesic use (Figure 1). The trajectory group of 'no use' comprised 691 (85.9%) women and was characterized by no or low probability of using analgesics during the study period (i.e., Year 1 to 10). The trajectory group of 'increasing probability to use', comprised 73 (9.1%) women. This group was characterized by a very low probability of analgesic use during the initial years (i.e., Year 1 to 3), followed by a steady increase in the probability of using analgesics. The trajectory group of 'constant analgesic use' comprised 40 (5.0%) women. This group was characterized by a high and constant probability of using analgesics during the study period. Table 1 lists the basic characteristics of study participants at Year 10.

There were 136 deaths recorded with a mean follow-up time of 9.6 years (standard deviation: 4.2), ranging from 1 to 15 years. Compared with the 'no use' group, the 'constant analgesic use' group was associated with 2.15 times higher risk of all-cause mortality (95% CI: 1.18 to 3.91) as shown in the multivariable Cox proportional hazard model. E-value results (Appendix S5) indicated that the observed HR estimate of 2.15 could be explained by an unmeasured confounder with the HR estimate of 2.78, but weaker confounding could not do so; the confidence interval could be moved to include the null by an unmeasured confounder with the HR estimate of 1.49, but weaker confounding could not do so. We did not find an association between 'increasing probability to use' versus 'no use' group (HR 0.84, 95% CI 0.46 to 1.54). When further adjustment was made for alcohol, physical activity, and occupation, the results were similar. Details can be found in Table 2. Of the 136 deaths observed, the most common cause was cancer (62, 45.6%), followed by other causes (42, 30.9%) and cardiovascular disease (32, 23.5%). No association between cause-specific mortality and pattern of analgesic use was found (Table 2).

A total of 626 women reported data on quality of life. Both the 'increasing probability to use' and the 'constant analgesic use' group were associated with worse quality of life across subscales of physical function, role limitations due to physical health and pain qual, compared with the 'no use' group (e.g., 'increasing probability to use' vs 'no use'; subscale of physical function; MD -23.5,

95% CI -30.1 to -16.9). These effect estimates were larger than the minimum clinical important difference (MCID) of 10 points (0-100 scale, SF-36) for the upper confidence limit of a negative value[19]. The 'constant analgesic use' group, rather than the 'increasing probability to use' group, was associated with worse quality of life in subscale of general health (MD -5.9, 95% CI -10.1 to -1.7). But the effect estimate was too small to have MCID. Both the 'increasing probability to use' and the 'constant analgesic use' group were not associated with worse quality of life across other remaining four subscales, compared with the 'no use' group (e.g., 'increasing probability to use' vs 'no use'; subscale of emotional well-being; MD -0.27, 95% CI -2.9 to 2.3). Details can be found in Figure 2, Figure 3 and Appendix S6.

Sensitivity results indicated that a slight increase in the risk estimate when comparing the redefined 'constant analgesic use' group which had one or more waves of opioids versus 'no use' group (HR 2.93, 95% CI 1.44 to 5.94 vs original HR 2.15, 95% CI 1.18 to 3.91). The results were similar when redefining 'constant analgesic use' group as the group with two or more waves of opioids (HR 2.83, 95% CI 1.13 to 7.10). However, the risk estimate decreased (HR 1.38, 95% CI 0.50 to 3.84) when women who used opioid were excluded from the 'constant analgesic use' group. The risk estimates were similar when redefining 'increasing probability to use' group as the group with one or more waves of opioids. The results were similar after excluding women with cancer. For the lag period analyses, and compared with the 'no use' group, the results for the 'constant analgesic use' and the 'increasing probability to use' groups were similar. Details of the sensitivity and lag period analyses can be found in Appendix S6 and S7.

Discussion

In a representative cohort of middle-aged, community-dwelling women in UK, we identified three distinct trajectories of analgesic use: 'increasing probability to use', 'no use', and 'constant analgesic use'. While most women in our study (86%) had no or low probability of using analgesics during the study period, for the small group of women with constant and high probability of analgesic use, a 2-fold increase in the risk of all-cause mortality was observed. No association between cause-specific mortality and pattern of analgesic use was found. Worse quality of life in terms of physical function, role limitations due to physical health and pain was associated with constant and high probability and increased probability of using analgesics.

Compared with traditional approaches (e.g., mixed effects models and generalized estimating equations), group-based trajectory analysis can classify people into distinct, mutually exclusive groups, which allows us to explore beyond the population average[15,25]. One previous study including participants in inpatient multidisciplinary musculoskeletal rehabilitation used group-based trajectory analysis and identified six groups[29]. Our results were similar to those presented in this study. 'Increasing probability to use' group (9%) of the current study corresponded to group 2 (10.6%) in the previous study, which demonstrated increasing use during the follow-up period of 9 years. Our 'No use' (86%) group corresponded to group 1 (14%) which showed a constant and low use of analgesics. Likewise, our 'Constant analgesic use' group (5%) was similar with the study's groups 5 and 6 (49%) which showed stable use during the follow-up period of 9 years. Group 4 (10.9%) included a trajectory that was specific to an intervention and therefore, was not comparable to our results. And finally, group 3 (15.4%) resembled the shape of the population average. We did not find the similar group in this study, which might be due to the sample size issue. Moreover, these component percentage differences might be due to the population differences (inpatient vs community-dwelling women).

The markedly higher all-cause mortality in the 'constant analgesic use' group compared with the 'no use' group could be mainly attributed to the use of opioid, as suggested in our sensitivity analyses. The effect of opioid use on all-cause mortality was not observed in other trajectory groups, possibly due to the limited sample size, which should be confirmed by future studies with a larger sample size. We used lag period analyses to assess whether these results were affected by induction period, which confirmed that the results were robust. Limited sample size might be the main reason for non-statistically significant associations for cause-specific mortality. For the comparison between 'constant analgesic use' or 'increasing probability to use' group and 'no use' group, the worse outcomes in the SF-36 subscales of physical function and pain were similar to a previous study among medical cannabis patients[30]. This cross-sectional study indicated that patients who used pain medication tended to report higher levels of pain and lower levels of physical component score from the Short Form-12 Health Survey (SF-12), but no significant difference was observed for mental component score from SF-12. Our cohort study confirmed this finding in a general population of women and further proposed that not only the 'constant analgesic

use' group but also the 'increasing probability to use' group would lead to poorer physical health. The impact of analgesic use on long-term mental health remains debatable, however. A previous meta-analysis has indicated that anti-inflammatory treatment might decrease depressive symptoms in adults who have either a diagnosis of depression or experience depressive symptoms[31]. It is possible, however, the relationship between analgesic use and mental disorders (e.g., depression) is bi-directional, future studies should consider using repeatedly measured data (exposures, covariates and outcomes) to correct for this potential bias[12].

Our study had several strengths. To our knowledge, this is the first study to identify the trajectories of analgesic use in the general women population, and then relate these trajectories to mortality and quality of life. The data we used, contain a long-term follow-up with good recruitment and retention rates. We performed additional analyses to confirm the robustness of our results: E-value for unmeasured confounding, lag period analysis for induction period and sensitivity analyses for the effect estimate from non-opioid and opioid analgesics.

Some limitations should also be mentioned. Firstly, our data did not have detailed information about the dosage and frequency of analgesic use, which prevented us from exploring the potential dose-response relationship. Overdose or inappropriate choice of analgesic should be explored in future studies. Moreover, Chingford 1000 Women Study included middle-aged UK women, and almost all women were white. We must exercise caution when generalising the results to men, other age groups, other ethnic groups, or to other countries. Thirdly, unmeasured confounding still might affect the results, although we adjusted extensively to several covariates. Our reported E-values, at least partially, supported the robustness of the results. Finally, confounding by indication might affect the results. Although we adjusted major illness or operations and performed a sensitivity analysis by excluding women with cancer, unrecorded disease and/or disease severity might have affected the analgesic prescription.

Conclusions

In this cohort of middle-aged women, a small group of women had a high probability of using analgesic and a markedly higher all-cause mortality compared with those with no or low probability of using analgesics.

References

1. Jackson T, Thomas S, Stabile V, Han X, Shotwell M, McQueen K. Prevalence of chronic pain in low-income and middle-income countries: a systematic review and meta-analysis. *Lancet (London, England).* 2015;385 Suppl 2:S10.

2. Dahlhamer J, Lucas J, Zelaya C, et al. Prevalence of Chronic Pain and High-Impact Chronic Pain Among Adults - United States, 2016. *MMWR Morbidity and mortality weekly report.* 2018;67(36):1001-1006.

3. McDonagh MS, Selph SS, Buckley DI, et al. AHRQ Comparative Effectiveness Reviews. *Nonopioid Pharmacologic Treatments for Chronic Pain*. Rockville (MD): Agency for Healthcare Research and Quality (US); 2020.

4. Megale RZ, Deveza LA, Blyth FM, et al. Efficacy and Safety of Oral and Transdermal Opioid Analgesics for Musculoskeletal Pain in Older Adults: A Systematic Review of Randomized, Placebo-Controlled Trials. *The journal of pain.* 2018;19(5):475.e471-475.e424.

5. Saragiotto BT, Machado GC, Ferreira ML, Pinheiro MB, Abdel Shaheed C, Maher CG. Paracetamol for low back pain. *The Cochrane database of systematic reviews.* 2016;2016(6):Cd012230.

6. Wilson N, Kariisa M, Seth P, Smith Ht, Davis NL. Drug and Opioid-Involved Overdose Deaths - United States, 2017-2018. *MMWR Morbidity and mortality weekly report.* 2020;69(11):290-297.

7. Jørgensen CC, Petersen M, Kehlet H, Aasvang EK. Analgesic consumption trajectories in 8975 patients 1 year after fast-track total hip or knee arthroplasty. *European journal of pain (London, England).* 2018.

8. Fredheim OM, Brelin S, Hjermstad MJ, et al. Prescriptions of analgesics during complete disease trajectories in patients who are diagnosed with and die from cancer within the five-year period 2005-2009. *European journal of pain (London, England).* 2017;21(3):530-540.

9. Riley RD, van der Windt D, Croft P, Moons KG. *Prognosis research in healthcare: concepts, methods, and impact.* Oxford University Press; 2019.

10. Chen T, Du J, Zhong N, et al. Trajectories of heroin use predict relapse risk among heroin-dependent patients: A 5-year follow-up study. *Journal of Clinical Neuroscience.* 2020;76:134-139.

11. Dublin S, Walker RL, Gray SL, et al. Prescription opioids and risk of dementia or cognitive decline: a prospective cohort study. *Journal of the American Geriatrics Society.* 2015;63(8):1519-1526.

12. Lash TL, VanderWeele TJ, Haneause S, Rothman K. *Modern epidemiology.* Lippincott Williams & Wilkins; 2020.

13. Arden NK, Griffiths GO, Hart DJ, Doyle DV, Spector TD. The association between osteoarthritis and osteoporotic fracture: the Chingford Study. *British journal of rheumatology.* 1996;35(12):1299-1304.

14. Rod NH, Bengtsson J, Budtz-Jørgensen E, et al. Trajectories of childhood adversity and mortality in early adulthood: a population-based cohort study. *Lancet (London, England).* 2020;396(10249):489-497.

15. Song M, Hu FB, Wu K, et al. Trajectory of body shape in early and middle life and all cause and cause specific mortality: results from two prospective US cohort studies. *BMJ (Clinical research ed).* 2016;353:i2195.

16. McHorney CA, Ware JE, Jr., Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Medical care.* 1993;31(3):247-263.

17. Laucis NC, Hays RD, Bhattacharyya T. Scoring the SF-36 in Orthopaedics: A Brief Guide. *The Journal of bone and joint surgery American volume.* 2015;97(19):1628-1634.

18. Ferreira ML, Herbert RD, Crowther MJ, Verhagen A, Sutton AJ. When is a further clinical trial justified? *BMJ (Clinical research ed).* 2012;345:e5913.

19. 36-Item Short Form Survey (SF-36) Scoring Instructions. https://www.rand.org/health-care/surveys_tools/mos/36-item-short-form/scoring.html. Accessed 20 Dec, 2020.

20. Global burden of 87 risk factors in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet (London, England).* 2020;396(10258):1223-1249.

21. Kluzek S, Sanchez-Santos MT, Leyland KM, et al. Painful knee but not hand osteoarthritis is an independent predictor of mortality over 23 years follow-up of a population-based cohort of middle-aged women. *Annals of the rheumatic diseases.* 2016;75(10):1749-1756.

22. Nagin DS, Odgers CL. Group-based trajectory modeling in clinical research. *Annual review of clinical psychology.* 2010;6:109-138.

23. Jones BL, Nagin DS. A Note on a Stata Plugin for Estimating Group-based Trajectory Models. *Sociological Methods & Research.* 2013;42(4):608-613.

24. Radojčić MR, Arden NK, Yang X, et al. Pain trajectory defines knee osteoarthritis subgroups: a prospective observational study. *Pain.* 2020;161(12):2841-2851.

25. Herle M, Micali N, Abdulkadir M, et al. Identifying typical trajectories in longitudinal data: modelling strategies and interpretations. *European journal of epidemiology.* 2020;35(3):205-222.

26. Therneau TM, Lumley T. Package 'survival'. *R Top Doc.* 2015;128(10):28-33.

27. Van Buuren S. *Flexible imputation of missing data.* CRC press; 2018.

28. VanderWeele TJ, Ding P. Sensitivity Analysis in Observational Research: Introducing the E-Value. *Annals of internal medicine.* 2017;167(4):268-274.

29. Saltychev M, Laimi K, Pentti J, Kivimäki M, Vahtera J. Use of pain medication before and after inpatient musculoskeletal rehabilitation: longitudinal analysis of a nationwide cohort. *International journal of rehabilitation research Internationale Zeitschrift fur Rehabilitationsforschung Revue internationale de recherches de readaptation.* 2018;41(2):159-165.

30. Perron BE, Bohnert K, Perone AK, Bonn-Miller MO, Ilgen M. Use of prescription pain medications among medical cannabis patients: comparisons of pain levels, functioning, and patterns of alcohol and other drug use. *Journal of studies on alcohol and drugs.* 2015;76(3):406-413.

31. Köhler O, Benros ME, Nordentoft M, et al. Effect of anti-inflammatory treatment on depression, depressive symptoms, and adverse effects: a systematic review and meta-analysis of randomized clinical trials. *JAMA psychiatry.* 2014;71(12):1381-1391.

Figure 1. Trajectories of analgesic use.

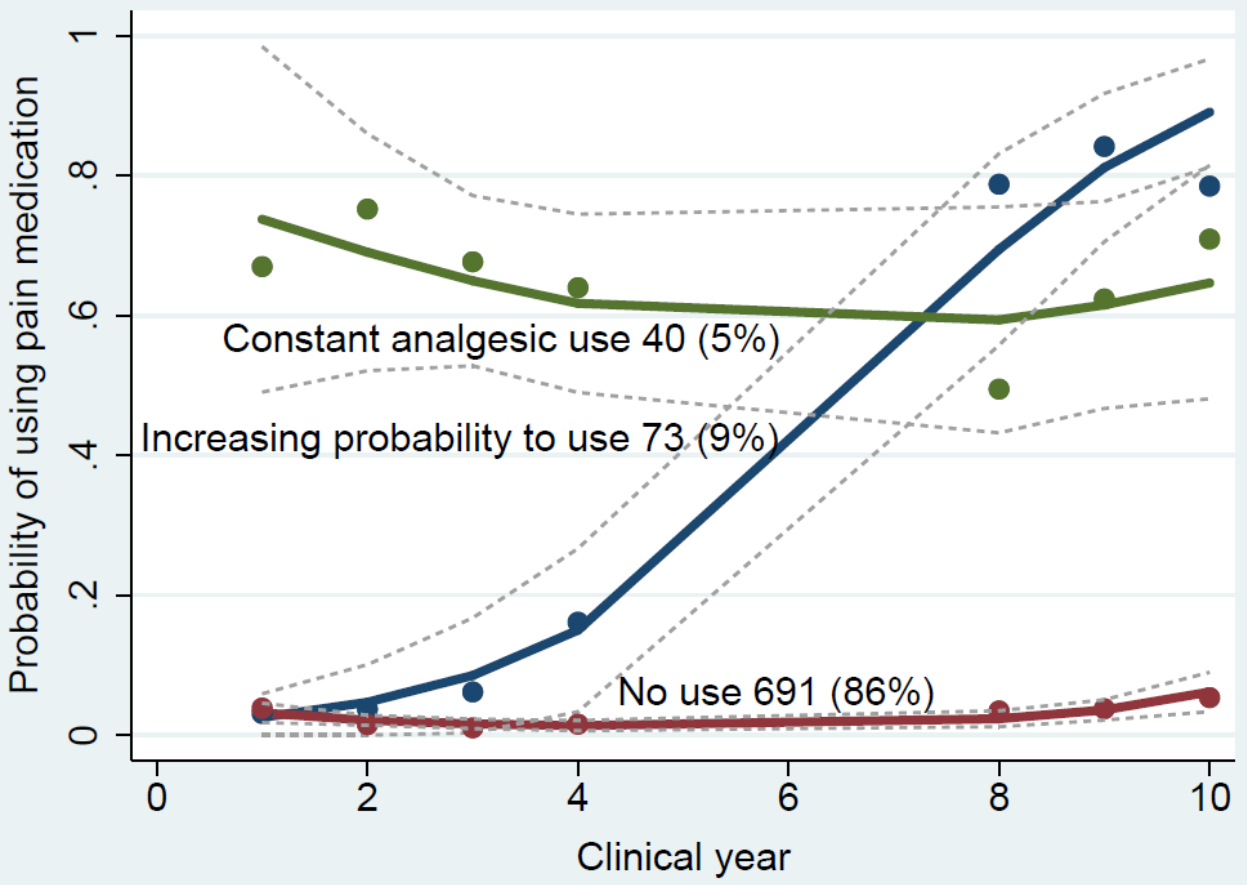
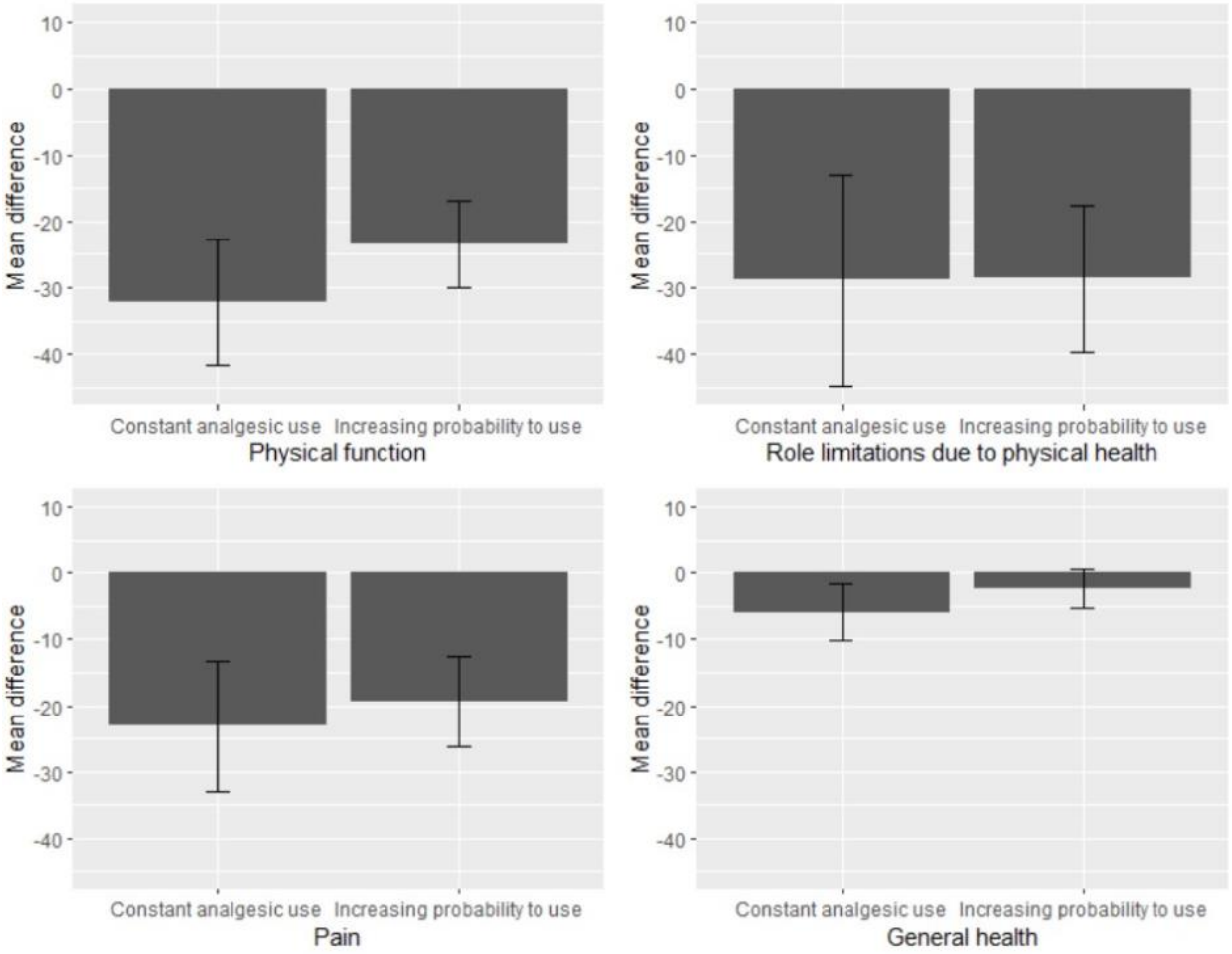
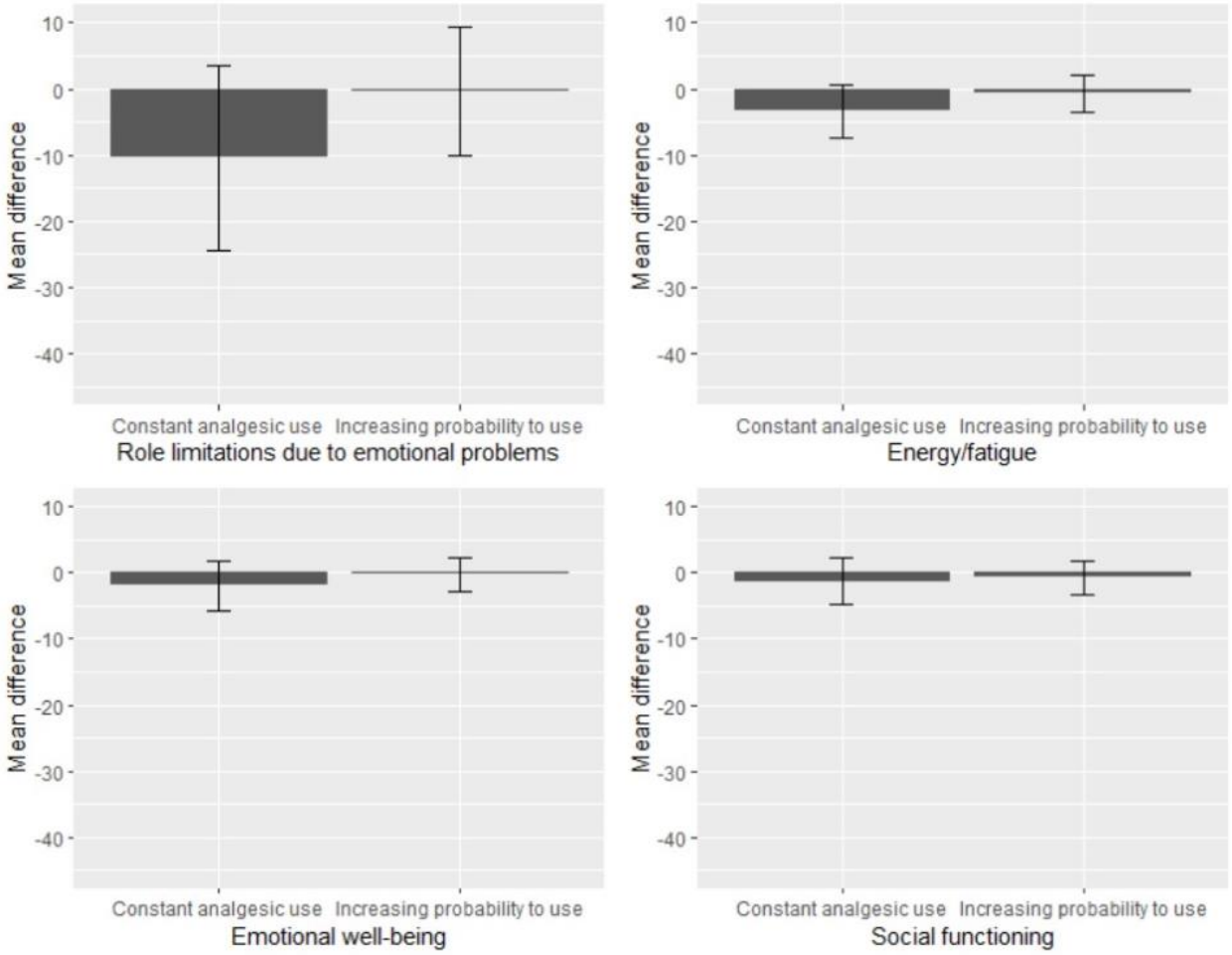


Figure 2. Quality of life results including four subscales related to physical component.



The reference group is ‘no use’ group. Error bars indicate 95% confidence interval.

Figure 3. Quality of life results including four subscales related to mental component.



The reference group is ‘no use’ group. Error bars indicate 95% confidence interval.

Table 1. Basic characteristics of study participants at Year 10 according to trajectories of analgesic use from Year 1 to 10.

Variable	Analgesic use trajectory			Whole cohort
	No use	Increasing probability to use	Constant analgesic use	
No (%) participants	691 (85.9)	73 (9.1)	40 (5.0)	804 (100.0)
Age	62.5 (5.9)	63.7 (6.1)	63.7 (5.7)	62.7 (5.9)
Body mass index	26.6 (4.6)	27.8 (5.0)	28.3 (5.4)	26.8 (4.7)
Missing (%)	1 (0.1)	N/A	N/A	1 (0.0.1)
Systolic blood pressure (mmHg)	136.1 (20.9)	136 (19.6)	133.2 (24.8)	135.9 (21.0)
Missing (%)	4 (0.6)	N/A	N/A	4 (0.5)
Smoking status (%)				
Never	390 (56.4)	34 (46.6)	19 (47.5)	443 (55.1)
Current	88 (12.7)	14 (19.2)	9 (22.5)	111 (13.8)
Ex-smoker	208 (30.1)	25 (34.2)	12 (30.0)	245 (30.5)
Missing	5 (0.7)	N/A	N/A	5 (0.6)
Fasting blood glucose (mmol/l)	5.2 (0.77)	5.3 (0.53)	5.3 (0.63)	5.3 (0.95)
Missing (%)	339 (49.1)	37 (50.7)	12 (30.0)	388 (48.3)
Cholesterol (mmol/l)	6.3 (1.2)	6.2 (1.1)	6.5 (1.2)	6.3 (1.2)
Missing (%)	334 (48.3)	36 (49.3)	11 (27.5)	381 (47.4)
Frequency of alcohol (%)				
Never	119 (17.2)	18 (24.7)	7 (17.5)	144 (17.9)
Weekly	265 (38.4)	30 (41.1)	16 (40.0)	311 (38.7)
Social occasions	307 (44.4)	25 (34.2)	17 (42.5)	349 (43.4)
Physical activity (%)				
Yes	543 (78.6)	65 (89.0)	31 (77.5)	639 (79.5)
No	138 (20.0)	8 (11.0)	8 (20.0)	154 (19.2)
Missing	10 (1.4)	N/A	1 (2.5)	11 (1.3)

Occupation (%)				
Manual	117 (16.9)	15 (20.5)	6 (15.0)	138 (17.2)
Non-manual	537 (77.7)	53 (72.6)	33 (82.5)	623 (77.5)
Missing	37 (5.4)	5 (6.8)	1 (2.5)	43 (5.3)
Major illness or operations (%)				
Yes	203 (29.4)	19 (26.0)	13 (32.5)	235 (29.2)
No	488 (70.6)	54 (74.0)	27 (67.5)	569 (70.8)

Values are means (SDs) unless stated otherwise.

Table 2. Association between trajectories of analgesic use and mortality (HR and 95% CI).

		Analgesic use trajectory	
Cause of death	**No use**	**Increasing probability to use**	**Constant analgesic use**
All-cause			
No of deaths (n=136)	111	12	13
Adjusted for age	1 (reference)	0.88 (0.49, 1.60)	2.30 (1.29, 4.09)
Multivariable adjusted[a]	1 (reference)	0.84 (0.46, 1.54)	2.15 (1.18, 3.91)
Further adjusted[b]	1 (reference)	0.78 (0.42, 1.44)	2.12 (1.16, 3.87)
Cancer			
No of deaths (n=62)	51	6	5
Adjusted for age	1 (reference)	1.01 (0.43, 2.36)	1.83 (0.73, 4.60)
Multivariable adjusted[a]	1 (reference)	0.99 (0.42, 2.34)	1.80 (0.71, 4.56)
Further adjusted[b]	1 (reference)	0.98 (0.41, 2.33)	1.87 (0.74, 4.77)
Cardiovascular disease			
No of deaths (n=32)	27	1	4
Adjusted for age	1 (reference)	0.29 (0.04, 2.11)	3.24 (1.13, 9.33)
Multivariable adjusted[a]	1 (reference)	0.27 (0.04, 2.00)	2.80 (0.82, 9.55)
Further adjusted[b]	1 (reference)	0.27 (0.04, 2.09)	2.39 (0.67, 8.50)
Other causes			
No of deaths (n=42)	33	5	4
Adjusted for age	1 (reference)	1.19 (0.46, 3.06)	2.49 (0.88, 7.07)
Multivariable adjusted[a]	1 (reference)	1.29 (0.46, 3.60)	2.54 (0.80, 8.01)
Further adjusted[b]	1 (reference)	1.32 (0.46, 3.80)	2.25 (0.69, 7.27)

HR: hazard ratio, CI: confidence interval

[a] age, systolic blood pressure, BMI, smoking status, glucose, cholesterol, major illness or operations,

[b] alcohol, physical activity, occupation

Appendix

Appendix S1. Details of data wave

Appendix S2. Details of analgesic

Appendix S3. Details of major illness or operations

Appendix S4. Methods for sensitivity analyses

Appendix S5. E-values for all-cause mortality

Appendix S6. Association between trajectories of analgesic use and quality of life (MD and 95% CI)

Appendix S7. Sensitivity and lag period analyses for all-cause mortality (HR and 95% CI)

Appendix S1. Details of data wave

No. reported waves	No. women	Accumulation no. women
Seven waves	300	300
Six waves	91	391
Five waves	154	545
Four waves	342	887
Three waves	68	955
Two waves	46	1001
One wave	2	1003

Appendix S2. Details of analgesic

With specific name

2.1.1 With opioid involved

Dihydrocodeine; Dextromoramide; Tramadol; Codeine; Morphine; Paracetamol and dextropropoxyphene; Paracetamol and Codeine; Paracetamol and dihydrocodeine

2.1.2 Without opioid involved

Indomethacin; Ibuprofen; Diclofenac; Etodolac; Fenbufen; Flurbiprofen; Fenoprofen; Mefenamic acid; Naproxen; Piroxicam; Ketoprofen; Movelat; Glucosamine; Feverfew; Paracetamol

2.2 Without specific name

NSAID; Anti-inflammatory; Analgesics; For migraine; Painkillers

Appendix S3. Details of major illness or operations

This covariate was defined by two categories of questionaries and the information from Year 1 to 10. If the participant has the serious operations/illnesses in at least one year, we defined the value of this covariate as yes. Otherwise, the value is no.

The first category is defined by Yes/No response. From Year 1 to Year 4, women were asked the following questions:

1) Serious operations/illnesses: cancer?

2) Serious operations/illnesses: Cardiac Vascular Diseases?

3) Serious operations/illnesses: Gastrectomy?

4) Serious operations/illnesses: Cholecystectomy?

The second category is defined by text response. From Year 1 to Year 4, women were asked the question: Serious operations/illnesses: Other? From Year 8 to Year 10, women were asked the question: Any major illnesses or operations?

The following texts in each year were included:

Year 1

heart valve ops
pyelitis
Rheumatic fever
Tuberculosis (and epilepsy since 14)
Non-A, non-B hepatitis and mild haemophiliac
Tuberculosis
Open heart surgery, rheumatic fever age 6-10
Rheumatic fever, heart murmur
appendectomy, also nervous breakdowns since the age of 19
Rheumatic fever x 2
Caesarians '70 Angina and cardiac spasms
Kidney stones, Asthma since '76, Appendicectomy '82, Dermatomyositis since '84 Psoriasis? Glandular fever and Hepatitis '85
Appendicectomy with Peritonitis
Heart valve replaced and appendicect. '56 and bladder rep. '87
Asthma
Diptheria
appendect,hepatitis.
appendix, tonsilectomy, Hiroshimo's thyroid.
Typhoid (as child)
sarcoidosis
SLE (Systemic Lupus Erythematosus)
Multiple sclerosis
RA (Rheumatoid Arthritis)
Polio
Typhoid
Epilepsy
Ulcerative colitis
Kidney removed
Granes Disease
angina, gout
Parkinsons & Maiges syndrome
oopherectomy
infectious hepatitis
TB
Psoriatic arthritis
ectopic(peritonitis)
gastric ulcer
Psitticosis
Rheumatic Fever, heart failure, kidney failure
RA

jaundice
Viral meningitis, ectopic (74),
Hepatitis
R eye blind; 5 ops
Viral meningitis
TB, pleuresy
rheumatic fever
SLE
Urethral repair
duodenal ulcer;
D.V.T. L leg, 1966 ect.pregnancy
osteomyelitis
viral hepatitis
Crohns
Rheumatic Fever, Polyarthritis
encephalitis
temp colostomy
glaucoma
polio head
L nephrectomy
Bovine TB,breast lump
ectopic; fall. Tubes removed age?
over active thyroid
preg toxaemia; breast lump '72; chemical imbalance +depression, '80
TB op
septicaemia myelitis
cerebral palsy; left
breast lump, asthma
breast lumps rem, stomach ulcer
cataracts, DM ops
thyroid tumour
epilepsy; partially blind
cholycystitis, congenital spine curvature
malignant melanoma 1991 1990 l breast lump aspirated 1990 thyroidectomy benign tumours1981 Whiplash RTA
breech birth, R.hemiplegia with muscle wastage.
depression 2yrs
L knee septic arthritis, stopped age 8yrs after flares in both knees

Year 2

glaucoma
ME
High blood calcium, under Barts for thyroid.

gynae op
Hashimotos thyroiditis
Pelvic infectn
D & C 06/91 haemorrhage
TB glands as child, Radium
T.B
Ortho ops
Endocarditis Po
Gynae ops
TB as child
colostomy
Deep vein thrombosis
gynae(repair)
suspected MI
Gynae ops
gynae & molt op
Polip C
Gynae,ov. cyst ops,ov.cyst
Gynae ops,D&C
Gynae ops,D&C
gynae ops.

Year 3

depression
Lupus flare
burst abcess in uterus- TAH + BSO
Asthma
Asthma

Year 4

heart bypass
pacemaker
skin cancer
heart bypass

Year 8

Lump removed R breast. Hepatitis.
Asthma attacks
Viral infection for 6/12.
Jaundice.

Immune system broke down.
Cateract.
Glaucoma.
Sub arachnoid haemorrhage.
Blocked arteries.
Angina
Corpal suspension.
Hepatitis
Urostomy
Suspected heart attack Aug 97.
Gout.
Bowel operation.
Bowel operation.
CA on leg.
Transfusions at Whipps Cross
Aortic valve disease.
Lung biopsy (Bronchiolitis obliterans organising pneumonia)
Cone biopsy, ok. Septicaemia 04/98.

Year 9

Gall bladder removed. Pancreatic cyst drained.
Mastectomy R side. Cateract operation.
Ulcerative colitis.
Angina, attends chest clinic.
Thyroidism, Nov 97.
Bronchiectasis, Feb 98.
Sinus wash out, Sep 97.
Glaucoma & cateract operations.
Heart attack, Mar 97.
Asthma.
Bells palsy, 03/97.
Acute asthmatic episode, Dec 97.
Revision THR, June 97. Pulmonary embolism, Jul 97. Haemorrhage, Aug 97. Revision THR, Feb 98. Internal haemorrhage, May 98.
L DVT, Aug 97.
Treated for Bells Palsy
05/98 bad pneumonia
Rectoseal operation, 1/98; asthma now diagnosed.
Depression 9/98.
High BP, enlarged heart to be referred to a specialist at WX
Angina 08/98
hip pain 9/98, stroke 11/98

Hospital admission for exacerbation of constructive pulmonary airway disease/emphasema 12/97
Repair of pelvic floor 03/98; seeing private psychiatrist for depression at Holly House
Admitted to WX, diagnosed hypothyroidism and anemia 8/98
Emphysema, 6/98
MS 20yrs thus effecting balance breast cyst remove 4/98
Stomach ulcer diagnosed 02/98
Attending pain clinic for cerebral palsy, since 12/98
Lung surgery 01/98
Lumpectomy & radiotherapy 11/98
L cataract 06/98, also attends psychiatric clinic
Psorosisrosia 12/98 , brain haemorrhage 01/99
2 heart attacks 1st one silent, 2nd in 02/99
Septicaemia 04/98

Year 10

Shunt placed between pancreas and stomach for pseudo cyst.
Chemotherapy for recurring NH lymphoma.
Ovarian cyst removed, chemo.
Angina
Chronic asthma
Stroke 09/98 hand and voice speech quite good
Angina
Heart attack 06/98, diagnosed angina
Breast Ca 12/98, 2 lumps removed
Septic Arthritis
Hysterectomy 02/98 cancerous polyp removed
Fibrillation 05/99
Revision of L THR, Jun 98; pulmonary embolism, Jul 97; haemorrhage, Aug 97; revision L THR, Feb 98; internal haemorrhage, May 98.
Stroke September 1998 no lasting affects
Another breast ca in same breast 04/99 masectomy
Numbness L side, OA hands, colitis
Womb being investigated due to periods on non-bleed HRT, seeing Dr 09/99
Brain operation April 1999
Diagnosed enlarged left ventricle July 1999
Smear abnormal cone biopsy arranged 01/00. Cataracts diagnosed 12/99.
Hysterectomy August 1999 Ca of ovaries July 1999
Resection of small intestine due to adhesions in September 1999 in Hospital for a month
12/99 knocked down by car badly bruised and shaken
Polycytraemia (too many red blood cells 12/98)
January 2000 diagnosed with breast cancer resulting in lumpectomy and lymph nodes removed

Appendix S4. Methods for sensitivity analyses

redefined ‘constant analgesic use’ group as the group with one or more waves of opioids;

redefined ‘constant analgesic use’ group as the group with two or more waves of opioids;

redefined ‘constant analgesic use’ group as the group without opioids;

redefined ‘increasing probability to use’ group as the group with one or more waves of opioids;

exclude women with cancer (for the outcome – all-cause mortality and quality of life).

Appendix S5. E-values for all-cause mortality

	Analgesic use trajectory		
	No use	**Increasing probability to use**	**Constant analgesic use**
No of deaths (n=136)	111	12	13
Adjusted for age	1 (reference)	1.53 (1)	2.94 (1.67)
Multivariable adjusted[a]	1 (reference)	1.67 (1)	2.78 (1.49)
Further adjusted[b]	1 (reference)	1.88 (1)	2.74 (1.45)

a age, systolic blood pressure, BMI, smoking status, glucose, cholesterol, major illness or operations,

b alcohol, physical activity, occupation

Appendix S6. Association between trajectories of analgesic use and quality of life (MD and 95% CI)

S6.1. Whole cohort

	Analgesic use trajectory		
Sub-scales	**No use**	**Increasing probability to use**	**Constant analgesic use**
Physical functioning			
Adjusted for age	0 (reference)	-25.13 (-32.13, -18.13)	-34.34 (-44.38, -24.30)
Multivariable adjusted[a]	0 (reference)	-23.47 (-30.07, -16.87)	-32.25 (-41.75, -22.76)
Further adjusted[b]	0 (reference)	-23.46 (-30.10, -16.81)	-32.23 (-41.76, -22.70)
Role limitations due to physical health			

Adjusted for age	0 (reference)	-30.74 (-41.98, -19.50)	-31.81 (-47.93, -15.69)
Multivariable adjusted[a]	0 (reference)	-28.65 (-39.72, -17.57)	-28.82 (-44.76, -12.88)
Further adjusted[b]	0 (reference)	-28.60 (-39.75, -17.45)	-28.43 (-44.42, -12.44)
Role limitations due to emotional problems			
Adjusted for age	0 (reference)	-1.31 (-11.01, 8.38)	-11.61 (-25.52, 2.30)
Multivariable adjusted[a]	0 (reference)	-0.42 (-10.12, 9.29)	-10.39 (-24.35, 3.57)
Further adjusted[b]	0 (reference)	0.09 (-9.67, 9.86)	-10.30 (-24.29, 3.70)
Energy/fatigue			
Adjusted for age	0 (reference)	-0.63 (-3.42, 2.16)	-3.26 (-7.27, 0.75)
Multivariable adjusted[a]	0 (reference)	-0.67 (-3.47, 2.13)	-3.30 (-7.33, 0.73)
Further adjusted[b]	0 (reference)	-0.68 (-3.50, 2.13)	-3.18 (-7.23, 0.86)
Emotional well-being			
Adjusted for age	0 (reference)	-0.63 (-3.42, 2.16)	-3.26 (-7.27, 0.75)
Multivariable adjusted[a]	0 (reference)	-0.27 (-2.86, 2.32)	-2.01 (-5.74, 1.71)
Further adjusted[b]	0 (reference)	-0.63 (-3.21, 1.96)	-1.82 (-5.52, 1.89)
Social functioning			
Adjusted for age	0 (reference)	-0.92 (-3.39, 1.55)	-1.38 (-4.92, 2.17)
Multivariable adjusted[a]	0 (reference)	-0.79 (-3.27, 1.69)	-1.27 (-4.84, 2.30)
Further adjusted[b]	0 (reference)	-0.71 (-3.21, 1.78)	-1.35 (-4.93, 2.23)
Pain			
Adjusted for age	0 (reference)	-20.74 (-27.65, -13.81)	-24.91 (-34.84, -14.98)
Multivariable adjusted[a]	0 (reference)	-19.48 (-26.33, -12.63)	-23.14 (-33.00, -13.29)
Further adjusted[b]	0 (reference)	-19.31 (-26.21, -12.42)	-23.26 (-33.15, -13.38)
General health			
Adjusted for age	0 (reference)	-2.52 (-5.44, 0.40)	-5.92 (-10.11, -1.73)
Multivariable adjusted[a]	0 (reference)	-2.41 (-5.32, 0.49)	-5.92 (-10.11, -1.74)
Further adjusted[b]	0 (reference)	-2.53 (-5.45, 0.38)	-6.14 (-10.32, -1.96)

MD: mean difference, CI: confidence interval

a age, systolic blood pressure, BMI, smoking status, glucose, cholesterol, major illness or operations,

b alcohol, physical activity, occupation

S6.2. Exclude women with cancer

		Analgesic use trajectory	
Sub-scales	**No use**	**Increasing probability to use**	**Constant analgesic use**
Physical functioning			
Adjusted for age	0 (reference)	-24.61 (-31.84, -17.38)	-33.56 (-43.85, -23.27)
Multivariable adjusted[a]	0 (reference)	-23.13 (-29.93, -16.32)	-31.51 (-41.24, -21.78)
Further adjusted[b]	0 (reference)	-23.10 (-29.95, -16.25)	-31.58 (-41.35, -21.81)
Role limitations due to physical health			
Adjusted for age	0 (reference)	-29.15 (-40.73, -17.57)	-30.53 (-47.01, -14.04)
Multivariable adjusted[a]	0 (reference)	-27.19 (-38.60, -15.78)	-27.56 (-43.85, -11.27)
Further adjusted[b]	0 (reference)	-27.19 (-38.66, -15.72)	-27.23 (-43.57, -10.88)
Role limitations due to emotional problems			
Adjusted for age	0 (reference)	1.15 (-8.83, 11.14)	-12.78 (-26.99, 1.44)
Multivariable adjusted[a]	0 (reference)	2.01 (-7.97, 11.99)	-11.46 (-25.73, 2.81)
Further adjusted[b]	0 (reference)	2.48 (-7.55, 12.51)	-11.43 (-25.75, 2.88)
Energy/fatigue			
Adjusted for age	0 (reference)	-0.50 (-3.39, 2.38)	-3.10 (-7.21, 1.00)
Multivariable adjusted[a]	0 (reference)	-0.56 (-3.46, 2.33)	-3.19 (-7.32, 0.94)
Further adjusted[b]	0 (reference)	-0.55 (-3.46, 2.36)	-3.07 (-7.21, 1.07)
Emotional well-being			
Adjusted for age	0 (reference)	-0.50 (-3.39, 2.38)	-3.10 (-7.21, 1.00)
Multivariable adjusted[a]	0 (reference)	-0.70 (-3.40, 1.99)	-2.70 (-6.56, 1.16)
Further adjusted[b]	0 (reference)	-1.03 (-3.71, 1.66)	-2.47 (-6.31, 1.36)
Social functioning			
Adjusted for age	0 (reference)	-0.31 (-2.82, 2.21)	-0.19 (-3.77, 3.39)
Multivariable adjusted[a]	0 (reference)	-0.13 (-2.65, 2.39)	-0.09 (-3.70, 3.51)
Further adjusted[b]	0 (reference)	-0.13 (-2.67, 2.41)	-0.13 (-3.75, 3.49)

Pain			
Adjusted for age	0 (reference)	-20.00 (-27.06, -12.95)	-25.23 (-35.28, -15.18)
Multivariable adjusted[a]	0 (reference)	-18.82 (-25.82, -11.81)	-23.72 (-33.72, -13.72)
Further adjusted[b]	0 (reference)	-18.67 (-25.71, -11.64)	-23.93 (-33.96, -13.90)
General health			
Adjusted for age	0 (reference)	-2.54 (-5.53, 0.45)	-6.17 (-10.43, -1.91)
Multivariable adjusted[a]	0 (reference)	-2.38 (-5.36, 0.61)	-5.98 (-10.24, -1.72)
Further adjusted[b]	0 (reference)	-2.48 (-5.46, 0.51)	-6.22 (-10.48, -1.97)

MD: mean difference, CI: confidence interval

a age, systolic blood pressure, BMI, smoking status, glucose, cholesterol, major illness or operations,

b alcohol, physical activity, occupation

Appendix S7. Sensitivity and lag period analyses for all-cause mortality (HR and 95% CI)

	Analgesic use trajectory		
	No use	**Increasing probability to use**	**Constant analgesic use**
Sensitivity analyses			
Constant analgesic use group with one or more waves of opioid			
No of death (n=132)	111	12	9
Adjusted for age	1 (reference)	0.89 (0.49, 1.61)	3.10 (1.57, 6.12)
Multivariable adjusted[a]	1 (reference)	0.87 (0.47, 1.60)	2.93 (1.44, 5.94)
Further adjusted[b]	1 (reference)	0.81 (0.44, 1.50)	2.80 (1.38, 5.72)
Constant analgesic use group with two or more waves of opioid			
No of death (n=128)	111	12	5
Adjusted for age	1 (reference)	0.89 (0.49, 1.61)	3.08 (1.25, 7.58)
Multivariable adjusted[a]	1 (reference)	0.87 (0.47, 1.59)	2.83 (1.13, 7.10)
Further adjusted[b]	1 (reference)	0.81 (0.44, 1.49)	2.87 (1.14, 7.24)
Constant analgesic use group without opioid			
No of death (n=127)	111	12	4

Adjusted for age	1 (reference)	0.89 (0.49, 1.61)	1.45 (0.53, 3.93)
Multivariable adjusted[a]	1 (reference)	0.87 (0.47, 1.60)	1.38 (0.50, 3.84)
Further adjusted[b]	1 (reference)	0.81 (0.44, 1.50)	1.39 (0.49, 3.91)
Increasing probability to use group with one or more waves of opioid			
No of death (n=127)	111	3	13
Adjusted for age	1 (reference)	1.00 (0.32, 3.14)	2.30 (1.29, 4.08)
Multivariable adjusted[a]	1 (reference)	1.05 (0.32, 3.42)	2.11 (1.16, 3.82)
Further adjusted[b]	1 (reference)	0.96 (0.30, 3.13)	2.07 (1.14, 3.78)
Exclude women with cancer			
No of death (n=124)	103	10	11
Adjusted for age	1 (reference)	0.78 (0.41, 1.50)	2.08 (1.11, 3.88)
Multivariable adjusted[a]	1 (reference)	0.74 (0.38, 1.43)	1.99 (1.04, 3.79)
Further adjusted[b]	1 (reference)	0.70 (0.36, 1.37)	1.93 (1.01, 3.71)
Lag period analyses			
Lag period: 3 years			
No of death (n=120)	99	10	11
Adjusted for age	1 (reference)	0.82 (0.43, 1.58)	2.28 (1.22, 4.26)
Multivariable adjusted[a]	1 (reference)	0.78 (0.40, 1.52)	2.16 (1.14, 4.08)
Further adjusted[b]	1 (reference)	0.73 (0.37, 1.41)	2.08 (1.09, 3.96)
Lag period: 5 years			
No of death (n=112)	91	10	11
Adjusted for age	1 (reference)	0.90 (0.47, 1.73)	2.54 (1.36, 4.75)
Multivariable adjusted[a]	1 (reference)	0.87 (0.45, 1.70)	2.44 (1.28, 4.66)
Further adjusted[b]	1 (reference)	0.80 (0.41, 1.58)	2.41 (1.26, 4.63)
Lag period: 7 years			
No of death (n=97)	78	10	9
Adjusted for age	1 (reference)	1.06 (0.55, 2.04)	2.48 (1.24, 4.96)
Multivariable adjusted[a]	1 (reference)	1.01 (0.52, 1.98)	2.36 (1.15, 4.82)
Further adjusted[b]	1 (reference)	0.93 (0.48, 1.83)	2.33 (1.13, 4.80)
Lag period: 10 years			

No of death (n=73)	58	8	7
Adjusted for age	1 (reference)	1.14 (0.54, 2.39)	2.81 (1.28, 6.19)
Multivariable adjusted[a]	1 (reference)	1.17 (0.55, 2.47)	3.09 (1.37, 6.97)
Further adjusted[b]	1 (reference)	1.06 (0.50, 2.25)	3.12 (1.38, 7.06)

HR: hazard ratio, CI: confidence interval

a age, systolic blood pressure, BMI, smoking status, glucose, cholesterol, major illness or operations,

b alcohol, physical activity, occupation

CHAPTER SEVEN

Association of Chronic Musculoskeletal Pain with Mortality Among Middle-Aged UK Participants: A Population-Based Cohort Study with Mediation Analysis

Chapter Seven has been published as:

Chen L., Ferreira M.L., Nassar N., Preen D., Hopper J.L., Li S., Bui M., Beckenkamp P.R., Shi B., Arden N.K., Ferreira P.R. Association of Chronic Musculoskeletal Pain with Mortality Among Middle-Aged UK Participants: A Population-Based Cohort Study with Mediation Analysis. EClinicalMedicine. 2021 Nov 14;42:101202.

Statement from co-authors confirming authorship contribution of the PhD candidate

The co-authors of the paper "Association of Chronic Musculoskeletal Pain with Mortality Among Middle-Aged UK Participants: A Population-Based Cohort Study with Mediation Analysis" confirm that Lingxiao Chen has made the following contributions:

Conception and design of the research
Analysis and interpretation of the findings
Writing of the manuscript and critical appraisal of the content

In addition to the statements above, in cases where I am not the corresponding author of a published item, permission to include the published material has been granted by the corresponding author.

Lingxiao Chen Date: 21 July 2021

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

Manuela Loureiro Ferreira Date: 21 July 2021

EClinicalMedicine 42 (2021) 101202

Contents lists available at ScienceDirect

EClinicalMedicine

journal homepage: https://www.journals.elsevier.com/eclinicalmedicine

Research paper

Association of chronic musculoskeletal pain with mortality among UK adults: A population-based cohort study with mediation analysis

Lingxiao Chen[a,*], Manuela L Ferreira[a], Natasha Nassar[b], David B Preen[c], John L Hopper[d], Shuai Li[d,e,f], Minh Bui[d], Paula R Beckenkamp[g], Baoyi Shi[h], Nigel K Arden[i], Paulo H Ferreira[g]

[a] Faculty of Medicine and Health, Institute of Bone and Joint Research, The Kolling Institute, Northern Clinical School, University of Sydney, Sydney, NSW 2064, Australia
[b] Faculty of Medicine and Health, Children's Hospital Westmead Clinical School, University of Sydney, Sydney, Australia
[c] School of Population and Global Health, The University of Western Australia, Perth, Australia
[d] Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Australia
[e] Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom
[f] Precision Medicine, School of Clinical Sciences At Monash Health, Monash University, Clayton, VIC, Australia
[g] School of Health Sciences, Faculty of Medicine and Health, University of Sydney, Sydney, Australia
[h] Department of Biostatistics, Mailman School of Public Health, Columbia University, NY, United States
[i] Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, United Kingdom

ARTICLE INFO

Article History:
Received 24 August 2021
Revised 22 October 2021
Accepted 29 October 2021
Available online 14 November 2021

ABSTRACT

Background: We aimed to quantify the association between chronic musculoskeletal pain and all-cause mortality, and to investigate the extent to which this association is mediated by physical activity, smoking status, alcohol consumption, and opioid use.

Methods: For this population-based cohort study, we used data from UK Biobank, UK between baseline visit (2006–2010) to 18th December 2020. We assessed the associations between chronic musculoskeletal pain and all-cause mortality using a Cox proportional hazards model. We performed causal mediation analyses to examine the proportion of the association between chronic musculoskeletal pain and all-cause mortality.

Findings: Of the 384,367 included participants, a total of 187,269 participants reported chronic musculoskeletal pain. Higher number of pain sites was associated with increased risk of all-cause mortality compared to having no pain (e.g., four sites vs no site of pain, Hazard Ratio [HR] 1.46, 95% Confidence Interval [CI] 1.35 to 1.57). The multiple mediator analyses showed that the mediating proportions of all four mediators ranged from 53.4% to 122.6%: among participants with two or more pain sites, the effect estimate reduced substantially, for example, HR reduced from 1.25 (95% CI: 1.21 to 1.30; two pain sites) to 1.07 (95% CI: 1.01 to 1.11; two pain sites).

Interpretation: We found that higher number of pain sites was associated with increased risk of all-cause mortality compared to having no pain, and at least half of the association of chronic musculoskeletal pain with increased all-cause mortality may be accounted for by four mediators.

Funding: Twins Research Australia.

1. Introduction

The global burden of chronic musculoskeletal pain is substantial, with a recent systematic review indicating a 26% prevalence of chronic musculoskeletal pain in the general adult population and 39% in those older than 65 years [1]. It is still debatable whether chronic musculoskeletal pain is associated with higher risk of mortality, due possibly to the definition of musculoskeletal pain used by previous research. For example, one Danish study addressing this question defined musculoskeletal pain as pain in the last 14 days [2]; whereas an American study defined "frequent persistent" back pain as back pain symptoms reported in the past 12 months "most of the time" or "constantly" both at baseline and first follow-up visit [3]. These definitions may not be accurately representative of a population with chronic musculoskeletal pain, because the term "chronic" is defined as pain duration of at least 3 months [4].

Another limitation in the available literature relates to selection of appropriate comparison groups. For example, the aforementioned Danish and American studies defined the comparison group as no

* Corresponding author.
E-mail address: lche4036@uni.sydney.edu.au (L. Chen).

https://doi.org/10.1016/j.eclinm.2021.101202

Research in context

Evidence before this study

The association between chronic musculoskeletal pain and all-cause mortality, and the extent to which this association is mediated by physical activity, smoking status, alcohol consumption, and opioid use are still unclear. Previous studies were limited by methodological limitations, including ill-defined musculoskeletal pain (e.g. pain in the last 14 days), and inappropriate comparison groups (e.g. no musculoskeletal pain within 14 days). Previous studies have not comprehensively assessed the role of lifestyle factors and certain medications (such as opioids) as possible mediators between chronic musculoskeletal pain and mortality.

Added value of this study

To our knowledge, this is the first large population-based study to comprehensively assess the association between chronic musculoskeletal pain (type of pain and number of pain sites) and mortality (all-cause and cause-specific mortality). Further, it is also the first study to document that the association is mediated by lifestyle factors and opioid use, individually and simultaneously. We found that higher number of pain sites was associated with increased risk of all-cause mortality compared to having no pain, and at least half of the association of chronic musculoskeletal pain with increased all-cause mortality may be accounted for by four mediators.

Implications of all the available evidence

This cohort study provides further evidence that higher number of pain sites was associated with increased risk of all-cause mortality compared to having no pain. However, additional evidence is needed to assess the influence from pain-related symptoms (e.g. numbness and itching) and pain duration.

musculoskeletal pain within 14 days, and no back pain, respectively [2,3]. Participants with non-chronic musculoskeletal pain as well as those with other types of pain (e.g., stomach or abdominal pain and pain all over the body) should be excluded from the comparison group because, failing to do so might result in the underestimation of the association between chronic musculoskeletal pain and health outcomes and/or mortality. Moreover, the co-occurrence of chronic musculoskeletal conditions is often ignored despite their great impact on the management of the index condition. For example, one previous study based on the general Dutch population indicated that more than half of the population with chronic musculoskeletal pain reported pain at two or more sites [5]. Thus, it is important to appropriately account for co-occurrence of multisite pain when assessing the association between chronic musculoskeletal pain and mortality. Finally, previous studies have not comprehensively assessed the role of lifestyle factors and certain medications (such as opioids) as possible mediators between chronic musculoskeletal pain and mortality [3,6]. Only one previous study, with a limited sample size (n = 6324), explored three lifestyle factors (smoking, alcohol consumption, and physical activity) individually [6]. Patients with chronic pain were more likely to smoke, be inactive and use opioid regularly [7–9]. Patients with chronic pain were less likely to drink alcohol, and this behaviour could be partly due to opioid use [10]. These modifiable factors are known to increase mortality risk [11–14]. Thus, it is important to identify to what extent the association between chronic musculoskeletal pain and mortality is mediated via lifestyle factors and opioid use, when occurring individually and co-currently.

In this large prospective cohort study of middle-aged UK participants, we aimed to quantify the association between chronic musculoskeletal pain and mortality. The potential mediating roles of physical activity, smoking status, alcohol consumption, and opioid use were also explored.

2. Methods

2.1. Data

This study used data from the UK Biobank which recruited approximately 500,000 people aged 40–69 years between 2006 and 2010, from 22 centres in the UK [15]. This study was restricted to a subset at the initial assessment (2006–2010): participants with chronic musculoskeletal pain (neck or shoulder pain, back pain, hip pain, and knee pain) represented the exposed group, and those without pain made up the comparison group. Participants who experienced headache, facial pain, stomach or abdominal pain or pain all over the body were excluded. Participants who experienced musculoskeletal pain in the last month but did not report they ever had chronic musculoskeletal pain were also excluded. Details of the UK Biobank can be found in the registry online protocol: http://www.ukbiobank.ac.uk. The North West Multi-centre Ethics Committee granted ethical approval to access data from the UK Biobank, and all participants provided written informed consent. We report this study based on the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement [16]. The study was conducted under UK Biobank project number 56,837.

2.2. Exposure

As our study focused on musculoskeletal pain given it is a major contributor to the global burden of disease [17]. Musculoskeletal pain was defined using the options in the UK Biobank touchscreen questionnaire (Category 100,048) which includes headache, facial pain, neck or shoulder pain, back pain, stomach or abdominal pain, hip pain, knee pain and pain all over the body. Headache, facial pain, pain all over the body, and stomach or abdominal pain were not included in defining the exposure as there was insufficient evidence these were related to musculoskeletal conditions and were therefore beyond the scope of this study. Thus, chronic musculoskeletal pain was defined by responses of participants to two questions: 1. "In the last month have you experienced any of the following that interfered with your usual activities?"; 2. "Have you had neck or shoulder pains/back pains/hip pains/knee pains for more than 3 months?". Participants who answered yes to both questions were defined as participants who had chronic musculoskeletal pain. Considering the co-occurrence of chronic musculoskeletal pain conditions, we have divided the exposure into two parts: 1. type of pain for those with one pain site as: neck or shoulder pain only, back pain only, hip pain only and knee pain only; 2. number of pain sites as: one, two, three, or four pain sites. As question 2 follows question 1, the comparison group was composed by those who answered 'none of the above' to question 1. This is because if participants indicated they did not experience back pain in the last month that interfered with their usual activities (question 1), they would not be asked question 2: "Have you had back pains for more than 3 months?".

2.3. Outcome

Follow-up was ascertained from baseline, i.e., initial assessment visit when chronic musculoskeletal pain was measured (2006–2010); and continued until death was confirmed via the death registry, the participant withdrew from the study, or until the end of the follow-up period on 18th December 2020, whichever came first. The primary outcome was all-cause mortality.

The secondary outcome was cause-specific mortality which was identified from underlying (primary) cause of death in the death registry. Based on clinical knowledge and the large sample available in the UK Biobank, cause-specific mortality was defined as cancer (International Classification of Diseases 10th edition [ICD-10] codes C00 to C97), cardiovascular disease (ICD-10 codes I05 to I89), mental and behavioural disorder (ICD-10 codes F00 to F89), respiratory system disease (ICD-10 codes J09 to J99), suicide (ICD-10 codes X60 to X84), nervous system disease (ICD-10 codes G00 to G99), endocrine, nutritional and metabolic disease (ICD-10 codes E00 to E90), digestive system disease (ICD-10 codes K20 to K93), musculoskeletal system and connective tissue disease (ICD-10 codes M00 to M90), genitourinary system disease (ICD-10 codes N00 to N98), falls (ICD-10 codes W00 to W19), and others (remaining ICD-10 codes). We followed the ICD-10 definitions of death causes of morbidity and mortality and examined the outcomes 'death due to mental and behavioural disorders' (i.e. Chapter V Mental and behavioural disorders) and 'suicide' (i.e. Chapter XX External), separately.

2.4. Mediators

Physical activity, smoking status, alcohol consumption, and opioid use were included as potential mediators. These measures were assessed at the initial visit (2006–2010). Based on one previous study, physical activity participation was assessed using the International Physical Activity Questionnaire (IPAQ) activity group (low, <10.0; moderate, 10.0–49.9 and high, >=50 excess metabolic equivalent (MET)-hours/week) [18]. We used the data from the IPAQ activity group (Data-Field 22,032). The calculation methods could be found from the previous study [18]. The data was generated as part of UKB Application ID 12,184. The mediator of physical activity was modelled as low vs moderate or high. Based on one previous study, alcohol consumption was measured as alcohol intake frequency (daily or almost daily, three or four times a week, once or twice a week, one to three times a month, special occasions only and never; regular referred to the first three categories) [19]. The mediator of alcohol consumption was modelled as regular vs special occasions or never. Smoking status was defined as 'never', 'previous smoking', and 'current smoking'. The mediator of smoking status was modelled as current smoking vs never or previous smoking. Opioid use was defined using the regular medication use question (detailed names and codes can be found in Appendix S1). This questionnaire (Category 20,003) contains data on any regular treatments taken weekly, monthly, etc. (without doses and formulations). The mediator of opioid use was modelled as yes vs no. Two types of multiple mediators were created: one focused on lifestyle factors including physical activity, smoking status, and alcohol consumption; and the other combining all four.

2.5. Covariates

To avoid potential overadjustment, factors known to be associated with both chronic musculoskeletal pain and mortality as well as those occurring before chronic musculoskeletal pain was reported, were included as confounders [20]. These variables included age, sex, ethnicity, and the Townsend deprivation index [21]. Age was defined as a continuous variable. Sex was defined as a binary variable (female vs male). Ethnicity was defined as an unordered categorical variable (white, black, Asian, mixed, and other). Townsend deprivation index was defined as a continuous variable. The Townsend deprivation index is a composite measure of deprivation based on unemployment, non-car ownership, non-home ownership, and household overcrowding; a negative value represents high socioeconomic status. Each participant is assigned a score corresponding to the output area in which their postcode is located.

2.6. Statistical analysis

Descriptive statistics were used to describe the baseline characteristics (e.g. number of pain site, pain type, race/ethnicity, age, sex, Townsend deprivation index, body mass index, smoking status, alcohol consumption, physical activity, opioid use, mental wellbeing, and comorbidity) among participants with chronic musculoskeletal pain and the comparison group. We examined the association between chronic musculoskeletal pain and all-cause mortality using Cox proportional hazards regression models [22]. We established a stepped modelling framework: step 1, unadjusted analyses; and step 2, analyses adjusted for age, sex, ethnicity, and the Townsend deprivation index. Results from model 2 are reported in the results section. Hazard ratios (HRs) with 95% confidence intervals (CIs) were calculated. We firstly checked the proportional hazards (pH) assumption through goodness-of-fit test (cox.zph function from survival package) [23]. If any significant result was found, we then graphically assessed pH assumption through log-log Kaplan Meier plot [24]. Overall, the pH assumption was met. Chronic musculoskeletal pain was modelled with the type of pain and number of pain sites as described above. For the analysis of the type of pain, the exposure was examined using unordered categorical variables. For the analysis on the number of pain sites, we treated the number of pain sites as an unordered categorical variable initially and then performed a trend analysis (the number of pain sites was treated as a continuous variable) [20]. Results of the trend analysis are presented in Table 3. Cause-specific mortality was modelled through multistate survival model with the calculation of transition probability to account for competing risk of death due to other causes [25]. The above-mentioned stepped modelling framework was used. Complete case analysis (i.e. excluding participants with missing data in any included variable) was used for the main analysis given the percentage of missing data was negligible (e.g., 0.1% for Townsend deprivation index) [26]. The strategy to handle missing data in causal mediation analysis is listed below.

We performed causal mediation analyses to examine the proportion of the association between chronic musculoskeletal pain and all-cause mortality mediated by physical activity, smoking status, alcohol consumption, and opioid use [27]. We assumed the existence of potential interactions between the exposure and the mediator; and used regression-based approaches which allowed for the existence of exposure-mediator interaction to estimate the total effect, total natural indirect effect (TNIE) and total natural direct effect (TNDE) [27]. The TNIE represented the effect of chronic musculoskeletal pain on all-cause mortality that could be explained by its association with the inclusion of the mediator/s in the model. The TNDE represented the effect of chronic musculoskeletal pain on all-cause mortality that was independent of the mediator. The proportion of the association by the mediator (TNIE/[TNDE + TNIE]) was estimated to quantify the magnitude of mediation. Considering the missing data issue in some mediators (19.6% for physical activity, 0.4% for smoking status and 0.08% for alcohol consumption), bootstrap with multiple imputation was used to obtain robust HRs with 95% CIs.

For the primary outcome, several exploratory and sensitivity analyses were performed to confirm the robustness of the results (details could be found in Appendix S2). We examined whether the association between the exposure and all-cause mortality differed by sex, age, BMI, ethnicity, or smoking status through testing of multiplicative interactions using WALD statistics [22]. We used lag period analysis (excluding events which occur within 3-month, 6-month, 1-year, 3-year, 5-year and 7-year after

enrolment) to verify any potential induction period (exposure status at a given time will correlate with a possible increase or decrease in disease only at some later time), inverse probability treatment weighting with covariates which might be considered as confounders (e.g. depression and anxiety) to identify potential model misspecification, excluding participants with cancer at baseline given these death might be less likely to be caused primarily by musculoskeletal pain, and e-value to test for unmeasured confounding [20]. All statistical analyses were performed with tidyverse, rms, hmisc, survival, etm, and CMAverse [28] packages in R, version 4.04 (R Group for Statistical Computing). LC and PHF had access to the data.

3. Role of the funding source

The funding source had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

4. Results

4.1. Summary

Of the 384,367 included participants (Fig. 1), 208,412 (54.2%) were women, and the mean (SD) age was 57 (8) years (Table 1). A total of

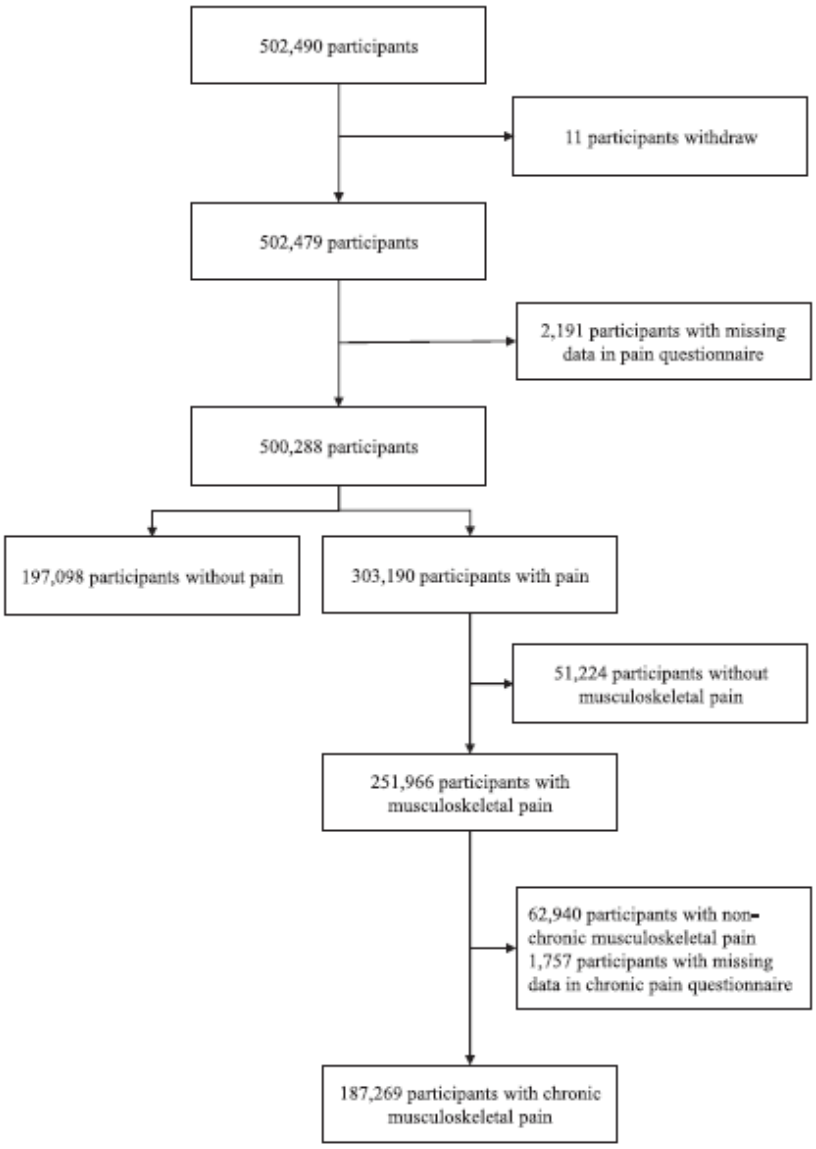


Fig. 1. Flow chart. The reasons for ineligibility and the numbers of ineligible participants were shown on the right arrow. The numbers of potential eligible participants were connected through the down arrow.

Table 1
Participant characteristics at the UK Biobank assessment.

Characteristic	No pain(n = 197,098)	Chronic pain(n = 187,269)	Total sample(n = 384,367)
Number of pain site			
One	NA	112,227 (59.9)	112,227 (29.2)
Two	NA	49,126 (26.2)	49,126 (12.8)
Three	NA	19,107 (10.2)	19,107 (5.0)
Four	NA	6809 (3.6)	6809 (1.8)
Pain type			
Neck or shoulder pain only	NA	31,331 (16.7)	31,331 (8.2)
Back pain only	NA	33,731 (18.0)	33,731 (8.8)
Hip pain only	NA	10,163 (5.4)	10,163 (2.6)
Knee pain only	NA	37,002 (19.8)	37,002 (9.6)
Mixed	NA	75,042 (40.1)	75,042 (19.5)
Race/ethnicity			
White	188,601 (95.7)	175,872 (93.9)	364,473 (94.8)
Black	2286 (1.2)	3221 (1.7)	5507 (1.4)
Asian	2707 (1.4)	4009 (2.1)	6716 (1.7)
Mixed	965 (0.5)	1169 (0.6)	2134 (0.6)
Other	2539 (1.3)	2998 (1.6)	5537 (1.4)
Age	56.7 (8.1)	57.2 (7.9)	57.0 (8.0)
Male	93,471 (47.4)	82,484 (44.1)	175,955 (45.8)
Townsend deprivation index, mean (SD)	−1.6 (2.9)	−1.1 (3.2)	−1.3 (3.1)
Missing	218 (0.1)	256 (0.1)	474 (0.1)
Body mass index, mean (SD)	26.7 (4.3)	28.3 (5.1)	27.5 (4.8)
Missing	781 (0.4)	1163 (0.6)	1944 (0.5)
Smoking status			
Current	17,401 (8.8)	22,776 (12.2)	40,177 (10.5)
Previous	65,866 (33.4)	69,266 (37.0)	135,132 (35.2)
Never	113,239 (57.5)	94,421 (50.4)	207,660 (54.0)
Missing	592 (0.3)	806 (0.4)	1398 (0.4)
Alcohol consumption			
Daily or almost daily	44,251 (22.5)	36,389 (19.4)	80,640 (21.0)
3–4 times a week	49,641 (25.2)	40,085 (21.4)	89,726 (23.3)
1–2 times a week	51,366 (26.1)	47,088 (25.1)	98,454 (25.6)
Unregular or never	51,725 (26.2)	63,517 (33.9)	115,242 (30.0)
Missing	115 (0.06)	190 (0.1)	305 (0.08)
Physical activity			
Low	27,662 (14.0)	30,101 (16.1)	57,763 (15.0)
Moderate	66,768 (33.9)	58,845 (31.4)	125,613 (32.7)
High	67,628 (34.3)	58,021 (31.0)	125,649 (32.7)
Missing	35,040 (17.8)	40,302 (21.5)	75,342 (19.6)
Opioid use	1771 (0.9)	18,811 (10.0)	20,582 (5.4)
Mental Wellbeing			
Depression	10,990 (5.6)	14,183 (7.6)	25,173 (6.5)
Anxiety	7417 (3.8)	9093 (4.9)	16,510 (4.3)
Comorbidity			
Diabetes	8599 (4.4)	12,044 (6.4)	20,643 (5.4)
Missing	324 (0.2)	673 (0.4)	997 (0.3)
Cancer	14,815 (7.5)	15,279 (8.2)	30,094 (7.8)
Missing	397 (0.2)	741 (0.4)	1138 (0.3)
Cardiovascular disease[a]			
One	47,047 (23.9)	55,778 (29.8)	102,825 (26.8)
Two	3811 (1.9)	7025 (3.8)	10,836 (2.8)
Three	778 (0.4)	1951 (1.0)	2729 (0.7)
Four	57 (0.0)	224 (0.1)	281 (0.1)

Data are presented as number (percentage) of patients unless otherwise indicated. For opioid use, depression and anxiety, participants who did not report them were treated as no use or no disease. For cardiovascular disease, we treated missing data as no disease to facilitate the calculation of the number of cardiovascular disease. We think it is fine because the percentage of missing data in the question - vascular/heart problems diagnosed by doctor (Data-Field 6150) is tiny (<0.3%).

[a] Four types of cardiovascular disease were included: heart attack, angina, stroke, high blood pressure.

187,269 participants reported chronic musculoskeletal pain: more than half reported one pain site (*n* = 112,227, 59.9%), followed by two (*n* = 49,126, 26.2%), three (*n* = 19,107, 10.2%) and four pain sites (6809, 3.6%). About one fifth reported knee pain only (*n* = 37,002, 19.8%), followed by back pain only (*n* = 33,731, 18.0%), neck or shoulder pain only (*n* = 31,331, 16.7%) and hip pain only (*n* = 10,163, 5.4%). Table 1 presents the participants' characteristics.

4.2. All-cause mortality

There were 25,917 deaths recorded over a mean follow-up time of 7.4 years (SD: 3.3, range: 0.0 to 14.6 years). Compared with participants without pain, the multivariable adjusted HR for all-cause mortality was 1.07 (95% CI 1.02 to 1.13) for participants with neck or shoulder pain only, 1.17 (95% CI 1.11 to 1.22) for participants with back pain only, 1.15 (95% CI 1.07 to 1.24) for participants with hip pain only and 1.03 (95% CI 0.99 to 1.08) for participants with knee pain only (Table 2). Participants with one (HR 1.09, 95% CI 1.06 to 1.12), two (HR 1.25, 95% CI 1.21 to 1.30), three (HR 1.43, 95% CI 1.36 to 1.51) and four (HR 1.46, 95% CI 1.35 to 1.57) pain sites had an increased risk of all-cause mortality (Table 2) compared to those without pain. Exploratory analyses (Appendix S3) indicated that for participants with two or more pain sites, younger age (<65 years) was associated with higher risk of all-cause mortality compared to

Table 2
Hazard Ratios (95% Confidence Intervals) for mortality according to pain type.

Cause of death	No pain(*n* = 197,098)	Neck or shoulder pain only (*n* = 31,331)	Back pain only(*n* = 33,731)	Hip pain only(*n* = 10,163)	Knee pain only(*n* = 37,002)
All cause					
No of deaths (*n* = 19,441)	11,877	1957	2305	769	2533
Risk of death,%	6.0	6.2	6.8	7.6	6.8
Unadjusted	1 (reference)	1.03 (0.98, 1.08)	1.14 (1.09, 1.19)	1.25 (1.16, 1.34)	1.14 (1.10, 1.19)
Multivariable adjusted[a]	1 (reference)	1.07 (1.02, 1.13)	1.17 (1.11, 1.22)	1.15 (1.07, 1.24)	1.03 (0.99, 1.08)
Cancer					
No of deaths (*n* = 10,363)	6502	1000	1172	388	1301
Risk of death,%	3.3	3.2	3.5	3.8	3.5
Unadjusted	1 (reference)	0.96 (0.90, 1.03)	1.05 (0.99, 1.12)	1.16 (1.05, 1.29)	1.07 (1.01, 1.14)
Multivariable adjusted[a]	1 (reference)	1.00 (0.94, 1.07)	1.10 (1.03, 1.17)	1.06 (0.96, 1.18)	0.99 (0.93, 1.05)
Endocrine, nutritional and metabolic disease					
No of deaths (*n* = 196)	109	34	26	8	19
Risk of death,%	0.06	0.1	0.08	0.08	0.05
Unadjusted	1 (reference)	1.95 (1.33, 2.86)	1.39 (0.91, 2.14)	1.44 (0.70, 2.94)	0.94 (0.58, 1.52)
Multivariable adjusted[a]	1 (reference)	1.99 (1.35, 2.93)	1.34 (0.87, 2.05)	1.39 (0.68, 2.86)	0.81 (0.50, 1.33)
Mental and behavioural disorder					
No of deaths (*n* = 410)	239	40	62	13	56
Risk of death,%	0.1	0.1	0.2	0.1	0.2
Unadjusted	1 (reference)	1.04 (0.74, 1.45)	1.51 (1.14, 2.00)	1.06 (0.61, 1.86)	1.26 (0.94, 1.69)
Multivariable adjusted[a]	1 (reference)	1.09 (0.78, 1.52)	1.58 (1.19, 2.09)	0.92 (0.52, 1.61)	1.09 (0.81, 1.46)
Nervous system disease					
No of deaths (*n* = 1048)	649	110	120	42	127
Risk of death,%	0.3	0.4	0.4	0.4	0.3
Unadjusted	1 (reference)	1.05 (0.86, 1.29)	1.08 (0.89, 1.31)	1.26 (0.93, 1.73)	1.05 (0.87, 1.27)
Multivariable adjusted[a]	1 (reference)	1.11 (0.91, 1.36)	1.16 (0.95, 1.40)	1.12 (0.82, 1.54)	0.96 (0.80, 1.16)
Cardiovascular disease					
No of deaths (*n* = 3728)	2245	377	436	163	507
Risk of death,%	1.1	1.2	1.3	1.6	1.4
Unadjusted	1 (reference)	1.05 (0.94, 1.17)	1.14 (1.02, 1.26)	1.42 (1.21, 1.66)	1.21 (1.10, 1.33)
Multivariable adjusted[a]	1 (reference)	1.12 (1.00, 1.25)	1.15 (1.04, 1.28)	1.35 (1.15, 1.58)	1.07 (0.97, 1.17)
Respiratory system disease					
No of deaths (*n* = 1290)	720	151	183	67	169
Risk of death,%	0.4	0.5	0.5	0.7	0.5
Unadjusted	1 (reference)	1.31 (1.10, 1.56)	1.49 (1.26, 1.75)	1.82 (1.42, 2.34)	1.26 (1.06, 1.49)
Multivariable adjusted[a]	1 (reference)	1.35 (1.13, 1.61)	1.49 (1.26, 1.75)	1.63 (1.27, 2.10)	1.08 (0.91, 1.28)
Digestive system disease					
No of deaths (*n* = 674)	382	74	84	21	113
Risk of death,%	0.2	0.2	0.2	0.2	0.3
Unadjusted	1 (reference)	1.21 (0.94, 1.55)	1.29 (1.02, 1.63)	1.07 (0.69, 1.66)	1.59 (1.29, 1.96)
Multivariable adjusted[a]	1 (reference)	1.26 (0.98, 1.62)	1.28 (1.01, 1.62)	1.03 (0.67, 1.60)	1.42 (1.15, 1.76)
Musculoskeletal system and connective tissue disease					
No of deaths (*n* = 64)	34	7	11	1	11
Risk of death,%	0.02	0.02	0.03	0.01	0.03
Unadjusted	1 (reference)	1.28 (0.57, 2.89)	1.89 (0.96, 3.73)	0.57 (0.08, 4.19)	1.73 (0.88, 3.42)
Multivariable adjusted[a]	1 (reference)	1.30 (0.57, 2.93)	1.89 (0.95, 3.74)	0.52 (0.07, 3.79)	1.55 (0.79, 3.07)
Genitourinary system disease					
No of deaths (*n* = 109)	62	14	14	4	15
Risk of death,%	0.03	0.04	0.04	0.04	0.04
Unadjusted	1 (reference)	1.41 (0.79, 2.52)	1.32 (0.74, 2.36)	1.26 (0.46, 3.46)	1.30 (0.74, 2.28)
Multivariable adjusted[a]	1 (reference)	1.39 (0.78, 2.49)	1.33 (0.74, 2.38)	1.09 (0.40, 3.01)	1.03 (0.58, 1.85)
Falls					
No of deaths (*n* = 138)	88	10	21	3	16
Risk of death,%	0.04	0.03	0.06	0.03	0.04
Unadjusted	1 (reference)	0.71 (0.37, 1.37)	1.40 (0.87, 2.25)	0.66 (0.21, 2.10)	0.97 (0.57, 1.66)
Multivariable adjusted[a]	1 (reference)	0.73 (0.38, 1.41)	1.41 (0.87, 2.27)	0.60 (0.19, 1.90)	0.84 (0.49, 1.43)
Suicide					
No of deaths (*n* = 155)	103	14	19	4	15
Risk of death,%	0.05	0.04	0.06	0.04	0.04
Unadjusted	1 (reference)	0.85 (0.49, 1.49)	1.08 (0.66, 1.76)	0.76 (0.28, 2.05)	0.78 (0.45, 1.34)
Multivariable adjusted[a]	1 (reference)	0.90 (0.51, 1.57)	1.00 (0.61, 1.63)	0.87 (0.32, 2.36)	0.77 (0.45, 1.32)
Other					
No of deaths (*n* = 1257)	744	126	157	46	184
Risk of death,%	0.4	0.4	0.5	0.5	0.5
Unadjusted	1 (reference)	1.05 (0.87, 1.27)	1.23 (1.04, 1.46)	1.21 (0.90, 1.63)	1.33 (1.13, 1.56)
Multivariable adjusted[a]	1 (reference)	1.07 (0.89, 1.30)	1.21 (1.02, 1.44)	1.14 (0.85, 1.54)	1.18 (1.00, 1.39)

[a] Adjusted for age, sex, townsend deprivation index and ethnicity.

older age with the same number of sites. The results from the sensitivity analyses (Appendix S4–S7) were similar to those of the main analyses (e.g. four pain sites vs no pain: original, HR 1.46, 95% CI 1.35 to 1.57; excluding participants who died within 3 months after enrolment, HR 1.46, 95% CI 1.35; excluding participants with cancer at baseline, HR 1.49, 95% CI 1.37 to 1.63).

4.3. *Cause-specific mortality*

The most common cause of death was cancer (13,488, 52.0%), followed by cardiovascular disease (5021, 19.4%) and respiratory system disease (1855, 7.2%) (Table 3). For participants with neck or shoulder pain only, there was a strong positive association with mortality

Table 3
Hazard Ratios (95% Confidence Intervals) for mortality according to number of pain sites.

Cause of death	No pain(n = 197,098)	One(n = 112,227)	Two(n = 49,126)	Three(n = 19,107)	Four(n = 6809)	Trend analysis(n = 384,367)
All cause						
No of deaths (n = 25,917)	11,877	7555	3949	1847	689	25,917
Risk of death,%	6.0	6.7	8.0	9.7	10.1	6.7
Unadjusted	1(reference)	1.12 (1.09, 1.15)	1.34 (1.29, 1.39)	1.62 (1.54, 1.70)	1.69 (1.57, 1.83)	1.16 (1.15, 1.17)
Multivariable adjusted[a]	1(reference)	1.09 (1.06, 1.12)	1.25 (1.21, 1.30)	1.43 (1.36, 1.51)	1.46 (1.35, 1.57)	1.12 (1.10, 1.13)
Cancer						
No of deaths (n = 13,488)	6502	3861	1975	859	291	13,488
Risk of death,%	3.3	3.4	4.0	4.5	4.3	3.5
Unadjusted	1(reference)	1.04 (1.00, 1.09)	1.22 (1.16, 1.29)	1.38 (1.28, 1.48)	1.31 (1.16, 1.47)	1.10 (1.08, 1.11)
Multivariable adjusted[a]	1(reference)	1.03 (0.99, 1.07)	1.15 (1.10, 1.21)	1.24 (1.15, 1.33)	1.15 (1.03, 1.30)	1.06 (1.04, 1.08)
Endocrine, nutritional and metabolic disease						
No of deaths (n = 295)	109	87	51	37	11	295
Risk of death,%	0.06	0.08	0.1	0.2	0.2	0.08
Unadjusted	1(reference)	1.40 (1.06, 1.86)	1.89 (1.36, 2.64)	3.55 (2.44, 5.15)	2.96 (1.59, 5.50)	1.42 (1.28, 1.56)
Multivariable adjusted[a]	1(reference)	1.32 (1.00, 1.75)	1.69 (1.21, 2.36)	2.91 (2.00, 4.27)	2.29 (1.23, 4.28)	1.33 (1.21, 1.47)
Mental and behavioural disorder						
No of deaths (n = 543)	239	171	77	43	13	543
Risk of death,%	0.1	0.2	0.2	0.2	0.2	0.1
Unadjusted	1(reference)	1.26 (1.03, 1.53)	1.30 (1.00, 1.68)	1.87 (1.35, 2.59)	1.59 (0.91, 2.77)	1.17 (1.08, 1.27)
Multivariable adjusted[a]	1(reference)	1.21 (0.99, 1.47)	1.17 (0.90, 1.52)	1.60 (1.16, 2.22)	1.35 (0.77, 2.36)	1.12 (1.03, 1.21)
Nervous system disease						
No of deaths (n = 1388)	649	399	202	102	36	1388
Risk of death,%	0.3	0.4	0.4	0.5	0.5	0.4
Unadjusted	1(reference)	1.08 (0.95, 1.22)	1.25 (1.07, 1.47)	1.64 (1.33, 2.02)	1.62 (1.16, 2.26)	1.14 (1.09, 1.20)
Multivariable adjusted[a]	1(reference)	1.07 (0.95, 1.22)	1.20 (1.02, 1.40)	1.51 (1.23, 1.87)	1.50 (1.07, 2.10)	1.12 (1.06, 1.18)
Cardiovascular disease						
No of deaths (n = 5021)	2245	1483	785	353	155	5021
Risk of death,%	1.1	1.3	1.6	1.8	2.3	1.3
Unadjusted	1(reference)	1.16 (1.09, 1.24)	1.41 (1.30, 1.53)	1.64 (1.46, 1.83)	2.01 (1.71, 2.37)	1.19 (1.15, 1.22)
Multivariable adjusted[a]	1(reference)	1.12 (1.05, 1.20)	1.31 (1.21, 1.43)	1.45 (1.29, 1.62)	1.73 (1.47, 2.04)	1.14 (1.11, 1.17)
Respiratory system disease						
No of deaths (n = 1855)	720	570	313	173	79	1855
Risk of death,%	0.4	0.5	0.6	0.9	1.2	0.5
Unadjusted	1(reference)	1.39 (1.25, 1.55)	1.75 (1.54, 2.00)	2.51 (2.12, 2.96)	3.21 (2.54, 4.05)	1.34 (1.29, 1.40)
Multivariable adjusted[a]	1(reference)	1.32 (1.18, 1.48)	1.57 (1.37, 1.79)	2.07 (1.75, 2.44)	2.50 (1.98, 3.16)	1.26 (1.21, 1.31)
Digestive system disease						
No of deaths (n = 967)	382	292	177	84	32	967
Risk of death,%	0.2	0.3	0.4	0.4	0.5	0.3
Unadjusted	1(reference)	1.34 (1.15, 1.56)	1.87 (1.56, 2.23)	2.29 (1.81, 2.90)	2.45 (1.71, 3.51)	1.31 (1.24, 1.38)
Multivariable adjusted[a]	1(reference)	1.29 (1.11, 1.51)	1.73 (1.44, 2.06)	1.98 (1.56, 2.51)	1.97 (1.37, 2.85)	1.25 (1.18, 1.32)
Musculoskeletal system and connective tissue disease						
No of deaths (n = 94)	34	30	17	11	2	94
Risk of death,%	0.02	0.03	0.03	0.06	0.03	0.02
Unadjusted	1(reference)	1.55 (0.95, 2.53)	2.01 (1.12, 3.60)	3.36 (1.70, 6.64)	1.71 (0.41, 7.13)	1.36 (1.14, 1.62)
Multivariable adjusted[a]	1(reference)	1.50 (0.92, 2.46)	1.84 (1.03, 3.31)	2.90 (1.46, 5.75)	1.42 (0.34, 5.92)	1.30 (1.09, 1.55)
Genitourinary system disease						
No of deaths (n = 159)	62	47	27	17	6	159
Risk of death,%	0.03	0.04	0.06	0.09	0.09	0.04
Unadjusted	1(reference)	1.33 (0.91, 1.95)	1.76 (1.12, 2.76)	2.86 (1.67, 4.89)	2.83 (1.22, 6.54)	1.35 (1.18, 1.55)
Multivariable adjusted[a]	1(reference)	1.22 (0.83, 1.79)	1.51 (0.96, 2.38)	2.24 (1.30, 3.84)	2.08 (0.89, 4.82)	1.25 (1.09, 1.44)
Falls						
No of deaths (n = 178)	88	50	28	7	5	178
Risk of death,%	0.04	0.04	0.06	0.04	0.07	0.05
Unadjusted	1(reference)	1.00 (0.71, 1.41)	1.28 (0.84, 1.96)	0.83 (0.38, 1.79)	1.66 (0.67, 4.08)	1.06 (0.92, 1.23)
Multivariable adjusted[a]	1(reference)	0.95 (0.67, 1.34)	1.15 (0.75, 1.76)	0.69 (0.32, 1.49)	1.31 (0.53, 3.24)	1.00 (0.87, 1.16)
Suicide						
No of deaths (n = 196)	103	52	22	11	8	196
Risk of death,%	0.05	0.05	0.04	0.06	0.1	0.05
Unadjusted	1(reference)	0.89 (0.64, 1.24)	0.86 (0.54, 1.37)	1.11 (0.60, 2.07)	2.27 (1.11, 4.66)	1.06 (0.92, 1.22)
Multivariable adjusted[a]	1(reference)	0.89 (0.64, 1.24)	0.90 (0.57, 1.43)	1.18 (0.63, 2.20)	2.47 (1.20, 5.10)	1.08 (0.94, 1.24)
Other						
No of deaths (n = 1733)	744	513	275	150	51	1733
Risk of death,%	0.4	0.5	0.6	0.8	0.7	0.5
Unadjusted	1(reference)	1.21 (1.08, 1.35)	1.49 (1.29, 1.71)	2.09 (1.76, 2.50)	1.99 (1.50, 2.64)	1.23 (1.18, 1.29)
Multivariable adjusted[a]	1(reference)	1.16 (1.04, 1.30)	1.36 (1.18, 1.56)	1.80 (1.51, 2.15)	1.65 (1.24, 2.20)	1.18 (1.13, 1.23)

[a] Adjusted for age, sex, townsend deprivation index and ethnicity.

related to endocrine, nutritional and metabolic disease (HR 1.99, 95% CI 1.35 to 2.93), and respiratory system disease (HR 1.35, 95% CI 1.13 to 1.61). Back pain only was associated with increased risk of mortality related to mental and behavioural disorder (HR 1.58, 95% CI 1.19 to 2.09), and respiratory system disease (HR 1.49, 95% CI 1.26 to 1.75). Mortality related to cardiovascular disease (HR 1.35, 95% CI 1.15 to 1.58), and respiratory system disease (HR 1.63, 95% CI 1.27 to 2.10) was more likely in participants with hip pain only, while people with knee pain only had greater risk of mortality from diseases of the digestive system (HR 1.42, 95% CI 1.15 to 1.76). Participants with two or more pain sites had a higher risk of death from respiratory system disease, and digestive system

disease, with hazard ratios ranging from 1.29 to 2.50. Appendix S8 lists transition probabilities (e.g. whole cohort; baseline to death due to cancer 3.5%; baseline to death due to cardiovascular disease 1.3%).

4.4. Mediation analyses

Table 4 presents the total, direct and indirect associations of chronic musculoskeletal pain with all-cause mortality as well as the proportion mediated. The single mediator analyses showed the following mediating proportions of the association between chronic musculoskeletal pain and all-cause mortality: 8.0 to 15.7% for physical activity; 32.5 to 79.0% for opioid use; 14.6 to 29.8% for smoking status and 2.4 to 17.5% for alcohol consumption. The multiple mediator analyses showed that, combined, the effect of the lifestyle factors ranged from 30.5 to 42.8%, and all four mediators ranged from 53.4 to 122.6%: for participants with one pain site, chronic musculoskeletal pain was not associated with all-cause mortality, for example, for

Table 4
Mediation analysis for the association between pain Type, number of pain sites, and all-cause mortality.

Mediators	Pain type			
	Neck or shoulder pain only	**Back pain only**	**Hip pain only**	**Knee pain only**
Total association[a]	**1.08 (1.05, 1.12)**	**1.17 (1.11, 1.22)**	**1.16 (1.08, 1.24)**	1.03 (0.98, 1.07)
Opioid use[b]				
Natural direct association	1.03 (0.97, 1.06)	**1.08 (1.03, 1.12)**	1.09 (0.99, 1.15)	0.98 (0.94, 1.03)
Natural indirect association	**1.03 (1.02, 1.05)**	**1.07 (1.05, 1.09)**	**1.05 (1.03, 1.08)**	**1.03 (1.02, 1.04)**
Proportion mediated,%	42.6	45.2	34.9	78.9
Smoking status[c]				
Natural direct association	**1.06 (1.01, 1.10)**	**1.13 (1.06, 1.18)**	**1.12 (1.05, 1.21)**	1.03 (0.98, 1.06)
Natural indirect association	**1.02 (1.02, 1.03)**	**1.04 (1.03, 1.04)**	**1.03 (1.02, 1.04)**	**1.01 (1.00, 1.01)**
Proportion mediated,%	29.8	23.9	23.3	19.0
Alcohol consumption[d]				
Natural direct association	**1.07 (1.01, 1.12)**	**1.15 (1.11, 1.23)**	**1.15 (1.09, 1.24)**	1.03 (0.98, 1.06)
Natural indirect association	**1.01 (1.01, 1.02)**	**1.01 (1.01, 1.01)**	**1.00 (1.00, 1.01)**	**1.00 (1.00, 1.01)**
Proportion mediated,%	17.5	8.2	2.4	15.4
Physical activity[e]				
Natural direct association	**1.08 (1.06, 1.11)**	**1.15 (1.10, 1.21)**	**1.13 (1.03, 1.21)**	1.03 (0.98, 1.08)
Natural indirect association	**1.01 (1.01, 1.01)**	**1.01 (1.01, 1.02)**	**1.01 (1.00, 1.02)**	**1.00 (1.00, 1.00)**
Proportion mediated,%	9.3	9.1	8.0	9.8
Lifestyle behaviours (smoking status, alcohol consumption, and physical activity)				
Natural direct association	**1.05 (1.02, 1.08)**	**1.11 (1.05, 1.17)**	**1.10 (1.01, 1.19)**	1.02 (0.97, 1.05)
Natural indirect association	**1.03 (1.03, 1.04)**	**1.06 (1.05, 1.07)**	**1.04 (1.03, 1.05)**	**1.01 (1.01, 1.02)**
Proportion mediated,%	41.0	38.0	30.5	41.8
All four				
Natural direct association	0.99 (0.97, 1.04)	1.04 (0.99, 1.08)	1.05 (0.97, 1.11)	0.97 (0.92, 1.02)
Natural indirect association	**1.07 (1.06, 1.08)**	**1.12 (1.11, 1.15)**	**1.09 (1.07, 1.12)**	**1.04 (1.03, 1.05)**
Proportion mediated,%	86.8	74.1	65.1	122.6
	Number of pain sites			
	One	**Two**	**Three**	**Four**
Total association	**1.09 (1.06, 1.12)**	**1.25 (1.21, 1.30)**	**1.43 (1.36, 1.51)**	**1.46 (1.35, 1.57)**
Opioid use				
Natural direct association	**1.03 (1.01, 1.06)**	**1.17 (1.14, 1.21)**	**1.30 (1.22, 1.36)**	**1.11 (1.03, 1.20)**
Natural indirect association	**1.04 (1.03, 1.05)**	**1.08 (1.07, 1.10)**	**1.11 (1.08, 1.14)**	**1.32 (1.24, 1.41)**
Proportion mediated,%	47.0	34.9	32.5	45.2
Smoking status				
Natural direct association	**1.07 (1.04, 1.11)**	**1.22 (1.18, 1.27)**	**1.38 (1.31, 1.46)**	**1.38 (1.29, 1.48)**
Natural indirect association	**1.02 (1.02, 1.03)**	**1.04 (1.03, 1.04)**	**1.05 (1.04, 1.06)**	**1.08 (1.06, 1.11)**
Proportion mediated,%	24.7	17.1	14.6	22.2
Alcohol consumption				
Natural direct association	**1.08 (1.06, 1.11)**	**1.23 (1.19, 1.26)**	**1.39 (1.35, 1.47)**	**1.40 (1.29, 1.48)**
Natural indirect association	**1.01 (1.01, 1.01)**	**1.02 (1.01, 1.02)**	**1.04 (1.03, 1.06)**	**1.04 (1.01, 1.08)**
Proportion mediated,%	10.5	8.7	13.0	12.7
Physical activity				
Natural direct association	**1.08 (1.06, 1.11)**	**1.23 (1.18, 1.27)**	**1.39 (1.32, 1.46)**	**1.40 (1.30, 1.52)**
Natural indirect association	**1.01 (1.01, 1.01)**	**1.01 (1.01, 1.02)**	**1.04 (1.03, 1.04)**	**1.05 (1.03, 1.08)**
Proportion mediated,%	9.1	9.8	11.8	15.7
Lifestyle behaviours (smoking status, alcohol consumption, and physical activity)				
Natural direct association	**1.06 (1.03, 1.09)**	**1.18 (1.14, 1.22)**	**1.31 (1.27, 1.36)**	**1.28 (1.20, 1.38)**
Natural indirect association	**1.04 (1.03, 1.04)**	**1.07 (1.06, 1.08)**	**1.12 (1.10, 1.13)**	**1.17 (1.13, 1.20)**
Proportion mediated,%	39.7	31.1	34.2	42.8
All four				
Natural direct association	1.01 (0.97, 1.04)	**1.07 (1.01, 1.11)**	**1.10 (1.03, 1.17)**	1.00 (0.92, 1.10)
Natural indirect association	**1.07 (1.06, 1.08)**	**1.13 (1.11, 1.14)**	**1.20 (1.16, 1.23)**	**1.30 (1.23, 1.39)**
Proportion mediated,%	73.7	54.1	53.4	69.6

Data are presented as hazard ratio (95% confidence interval) unless otherwise indicated. Effect estimates with statistical significance are labelled in bold.

[a] The effect estimate with its 95% confidence interval was slightly different for each mediation analysis. To reduce overlap, we listed the value from Tables 2 and 3.

[b] The mediator of opioid use was modelled as yes vs no.

[c] The mediator of smoking status was modelled as current smoking vs never or previous smoking.

[d] The mediator of alcohol consumption was modelled as regular vs special occasions or never.

[e] The mediator of physical activity was modelled as low vs moderate or high.

people with back pain only, HR 1.04, 95% CI 0.99 to 1.08 vs original, HR 1.17, 95% CI 1.11 to 1.22; for participants with two or more pain sites, the effect estimate presented a greater reduction, for example, HR reduced from 1.25 (95% CI 1.21 to 1.30; two pain sites) to 1.07 (95% CI 1.01 to 1.11; two pain sites).

5. Discussion

In a large population of middle-aged UK participants, neck or shoulder pain only, back pain only and knee pain only were associated with increased risk of all-cause mortality. Participants with higher number of pain sites had increased risk of all-cause mortality. However, these associations were mediated by physical activity, smoking status, alcohol consumption, and opioid use. At least half of the association of chronic musculoskeletal pain with increased all-cause mortality may be accounted for by four mediators.

The results of our all-cause mortality analyses are consistent with those of previous studies [3,6]. In contrast to previous studies, we clearly defined chronic musculoskeletal pain using two specific questions in UK Biobank. In addition, we were able to choose an appropriate comparison group by excluding participants with other types of pain and non-chronic musculoskeletal pain and assess the influence from the type of pain by focusing on those with one pain site. Considering number of pain sites, our results were similar to those of a recent Danish study [2], however the definition of musculoskeletal pain differs (ours: chronic vs the Danish study: pain in the last 14 days). Thus, the similar results might indicate a higher mortality risk for those with pain at multiple sites, compared to people with pain in one site, irrespective of the duration of pain. For participants experiencing pain in two or more sites, the results from our exploratory analysis which indicated that younger participants had higher risk of all-cause mortality, might reflect immortal time bias, as older participants might have been healthier enough to live longer at study entry compared to younger participants [29]. Considering cause-specific mortality analyses, our results are in line with previous research [2,3]. However, previous studies had smaller samples, yielding less precise estimates; they also did not consider the competing risk in analysing cause-specific mortality, which might have biased results [2,3,30]. We have a large sample size and used a multistate survival model, which allowed us to confirm that the association between respiratory system disease mortality or digestive system disease mortality was stronger than that of all-cause mortality among participants with two or more pain sites [25]. This might indicate that better management is needed for respiratory system disease or digestive system disease among participants with chronic musculoskeletal pain in two or more sites.

For the mediator - physical activity and alcohol consumption, our results are consistent with a previous study [6]. For the mediator - smoking status, our results showed that the relationship between chronic musculoskeletal pain and all-cause mortality was mediated by smoking status, which is in contrast to previous findings [6]. This difference might be due to the small sample size in the previous study (6324 vs ours: 384,367), which makes it difficult to detect moderate associations [6]. The results from our multiple mediator analyses indicated that most of the association between chronic musculoskeletal pain and all-cause mortality were mediated by all four factors, which means that poor lifestyle factors and opioid use may be the main drivers that increased the risk of all-cause mortality, rather than chronic musculoskeletal pain. Current guidelines indicate that: 1. people with chronic pain should remain physically active; 2. opioids should be avoided in general and only be used when the benefits outweigh the potential risks if other options fail [31–33]. Our study contributed to the field by providing more comprehensive and accurate results. Additionally, the results indicated that smoking cessation and alcohol consumption control should be added in future guidelines. Although educating patients to use less opioids is important, it can be challenging. Meanwhile, some people might need to use opioid if other therapies fail. One guideline mention that opioids should always be combined with nonpharmacologic and often nonopioid pharmacologic therapy [33], however, no specific guidance is provided. Our results showed that keeping adequate levels of physical activity, smoking cessation and alcohol consumption controlled should be emphasised to people who take opioids, as these healthy lifestyle behaviours could substantially decrease mortality risk.

To our knowledge, this is the first large population-based study to comprehensively assess the association between chronic musculoskeletal pain (type of pain and number of pain sites) and mortality (all-cause and cause-specific mortality). Further, it is also the first study to document that the association is mediated by lifestyle factors and opioid use, individually and simultaneously. Several additional analyses were performed to confirm the robustness of the results.

Some limitations need consideration. First, data on pain intensity and pain-related symptoms (e.g., numbness and itching) were not included in this study. Likewise, we did not have any data on the actual duration of symptoms and, therefore, could not ascertain the role of pain duration on the association between chronic musculoskeletal pain and mortality. Although there is no evidence that these factors could bias our results, future studies should untangle their roles. Second, the UK Biobank collected data from UK participants with specific ages (40–69), and we must exercise caution when generalising these findings to other age groups or other countries. Third, number of events in some categories for cause-specific mortality (e.g. falls and suicide) may be too small, and we caution the reader in making inferences based on these imprecise results. Fourth, the dose and duration of opioid use was not included so that overdose death could not be assessed. We defined opioid use through regular treatments taken weekly, monthly, etc. (short-term use was not included). Future studies should include data with doses, formulations and prescription dates (e.g. primary care data) to provide more accurate results. Fifth, due to the study scope, other types of chronic pain (e.g. stomach or abdominal pain) were not included. Assuming the definition of pain in the UK Biobank included eight different presentations (i.e. headache, facial pain, neck or shoulder pain, back pain, stomach or abdominal pain, hip pain, knee pain and pain all over the body), participants would have a total of 255 ($C_8^1 + C_8^2 + C_8^3 + C_8^4 + C_8^5 + C_8^6 + C_8^7 + C_8^8$) possible combinations considering pain status. Our analyses would be arguably underpowered if we were to include all the possible combinations. Sixth, insufficient primary care data in the current study makes it difficult to identify accurate cases of specific musculoskeletal pain diagnoses (e.g. autoimmune/rheumatic diseases) which could provide new insights in understanding the association between chronic musculoskeletal pain and mortality. The UK Biobank plans to release primary care data for all participants in future, providing the opportunity to further explore this issue. Seventh, we acknowledge certain limitations in the measurement of alcohol consumption and physical activity (ie. subjective measurement of alcohol consumption and physical activity; and qualitative measurement of alcohol consumption) could have biased the results. Objective and more comprehensive measurement should be explored in future studies. Finally, missing data in the mediators might have affected the results despite our approach of multiple imputation to handle this issue.

Higher number of pain sites was associated with increased risk of all-cause mortality compared to having no pain, and at least half of the association of chronic musculoskeletal pain with increased all-cause mortality may be accounted for by four mediators. Supporting healthy lifestyle behaviour (keeping adequate levels of physical activity, smoking cessation and alcohol consumption controlled) as well as opioids deprescription is an important strategy to decrease the mortality risk associated with chronic musculoskeletal pain.

Funding

Twins Research Australia.

Authors' contributions

All authors designed the study. LC conducted the data analysis and drafted the manuscript. All authors critically revised the manuscript for important intellectual content. All authors approved the final version of the manuscript. LC and PHF verified the underlying data reported in the manuscript. The corresponding author attests that all the listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Data sharing statement

Data from UK Biobank are available on application at www.ukbiobank.ac.uk/register-apply.

Declaration of Competing Interest

Dr Arden reported receiving personal fees from Pfizer/Lilly and Bristows LLP and grants from Merck outside the submitted work. Dr Nassar is supported by Financial Markets for Children and NHMRC. No other disclosures were reported.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.eclinm.2021.101202.

References

[1] Jackson T, Thomas S, Stabile V, Han X, Shotwell M, McQueen K. Prevalence of chronic pain in low-income and middle-income countries: a systematic review and meta-analysis. Lancet 2015;385(Suppl 2):S10.
[2] Holmberg T, Davidsen M, Thygesen LC, Krøll MJ, Tolstrup JS. Mortality among persons experiencing musculoskeletal pain: a prospective study among Danish men and women. BMC Musculoskelet Disord 2020;21(1):666.
[3] Roseen EJ, LaValley MP, Li S, Saper RB, Felson DT, Fredman L. Association of back pain with all-cause and cause-specific mortality among older women: a cohort study. J Gen Intern Med 2019;34(1):90–7.
[4] Nicholas M, Vlaeyen JW, Rief W, et al. The IASP classification of chronic pain for ICD-11: chronic primary pain. Pain 2019;160(1):28–37.
[5] Picavet HS, Schouten JS. Musculoskeletal pain in the Netherlands: prevalences, consequences and risk groups, the DMC(3)-study. Pain 2003;102(1–2):167–78.
[6] Smith D, Wilkie R, Croft P, Parmar S, McBeth J. Pain and mortality: mechanisms for a relationship. Pain 2018;159(6):1112–8.
[7] Boudreau D, Von Korff M, Rutter CM, et al. Trends in long-term opioid therapy for chronic non-cancer pain. Pharmacoepidemiol Drug Saf 2009;18(12):1166–75.
[8] Palmer KT, Syddall H, Cooper C, Coggon D. Smoking and musculoskeletal disorders: findings from a British national survey. Ann Rheum Dis 2003;62(1):33–6.
[9] van den Berg-Emons RJ, Schasfoort FC, de Vos LA, Bussmann JB, Stam HJ. Impact of chronic pain on everyday physical activity. Eur J Pain 2007;11(5):587–93.
[10] Ekholm O, Grønbaek M, Peuckmann V, Sjøgren P. Alcohol and smoking behavior in chronic pain patients: the role of opioids. Eur J Pain 2009;13(6):606–12.
[11] Macfarlane GJ, Beasley M, Jones GT, Stannard C. The epidemiology of regular opioid use and its association with mortality: prospective cohort study of 466 486 UK biobank participants. EClinicalMedicine 2020;21:100321.
[12] Smyth A, Teo KK, Rangarajan S, et al. Alcohol consumption and cardiovascular disease, cancer, injury, admission to hospital, and mortality: a prospective cohort study. Lancet 2015;386(10007):1945–54.
[13] Arem H, Moore SC, Patel A, et al. leisure time physical activity and mortality: a detailed pooled analysis of the dose-response relationship. JAMA Intern Med 2015;175(6):959–67.
[14] Jacobs DR, Adachi H, Mulder I, et al. Cigarette smoking and mortality risk: twenty-five-year follow-up of the seven countries study. Arch Intern Med 1999;159(7):733–40.
[15] Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. Plos Med 2015;12(3):e1001779.
[16] Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. Ann Intern Med 2007;147(8):573–7.
[17] Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the global burden of disease study 2019. Lancet 2020;396(10258):1204–22.
[18] Cassidy S, Chau JY, Catt M, Bauman A, Trenell MI. Cross-sectional study of diet, physical activity, television viewing and sleep duration in 233,110 adults from the UK Biobank; the behavioural phenotype of cardiovascular disease and type 2 diabetes. BMJ Open 2016;6(3):e010038.
[19] Jani BD, McQueenie R, Nicholl BI, et al. Association between patterns of alcohol consumption (beverage type, frequency and consumption with food) and risk of adverse health outcomes: a prospective cohort study. BMC Med 2021;19(1):8.
[20] Lash TL, VanderWeele TJ, Haneause S, Rothman K. Modern epidemiology. Lippincott Williams & Wilkins; 2020.
[21] Townsend P. Deprivation. J Soc Policy 1987;16(2):125–46.
[22] Harrell FE. Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis. Springer; 2015.
[23] Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. Biometrika 1994;81(3):515–26.
[24] Therneau TM, Lumley T. Package "survival". Survival analysis Published on CRAN 2014;2(3):119.
[25] Andersen PK, Abildstrom SZ, Rosthøj S. Competing risks as a multi-state model. Stat Methods Med Res 2002;11(2):203–15.
[26] Lee KJ, Tilling KM, Cornish RP, et al. Framework for the treatment and reporting of missing data in observational studies: The Treatment And Reporting of Missing data in Observational Studies framework. J Clin Epidemiol 2021;134:79–88.
[27] VanderWeele T. Explanation in causal inference: methods for mediation and interaction. Oxford University Press; 2015.
[28] Shi B, Choirat C, Coull BA, VanderWeele TJ, Valeri L. CMAverse: a suite of functions for reproducible causal mediation analyses. Epidemiology 2021;32(5):e20–2.
[29] Lévesque LE, Hanley JA, Kezouh A, Suissa S. Problem of immortal time bias in cohort studies: example using statins for preventing progression of diabetes. BMJ 2010;340:b5087.
[30] Lau B, Cole SR, Gange SJ. Competing risk regression models for epidemiologic data. Am J Epidemiol 2009;170(2):244–56.
[31] Chronic pain (primary and secondary) in over 16s: assessment of all chronic pain and management of chronic primary pain. London: National Institute for Health and Care Excellence (UK); April 7, 2021.
[32] Tauben D, Fishman S, Crowley M. Approach to the management of chronic non-cancer pain in adults. UpToDate. https://www.uptodate.com/contents/approach-to-the-management-of-chronic-non-cancer-pain-in-adults. 2021.
[33] Tauben D, Fishman S, Crowley M. Pharmacologic management of chronic non-cancer pain in adults. UpToDate. https://www.uptodate.com/contents/pharmacologic-management-of-chronic-non-cancer-pain-in-adults. 2021.

Appendix

Appendix S1. Details for opioids.

Appendix S2. Methods for exploratory and sensitivity analyses.

Appendix S3. Results for exploratory analyses.

Appendix S4. Results for sensitivity analyses: induction period.

Appendix S5. Results for sensitivity analyses: other covariates.

Appendix S6. Results for sensitivity analyses: exclude cancer patients.

Appendix S7. Results for sensitivity analyses: E-value.

Appendix S8. Transition probabilities.

Appendix S1. Details for opioids.

Name	Code
Tramadol	1140864992
Paracetamol + Tramadol	1141190956
Codeine	1140884444
Dihydrocodeine	1140884464
Medocodeine Tablet	1140856406
Ibuprofen + Codeine Phosphate	1140878030
Aspirin + Codeine 300mg/8mg tablet	1140882268
Aspirin + Codeine	1140882392
Pracetamol + Codeine	1140882394
Paracetamol + Dihydrocodeine tartrate	1140882396
Codeine phosphate + Kaolin 10mg/3g/10ml mixture	1140865654
Dihydrocodeine	1140884464
Morphine	1140871692
kaolin+morphine	1140882114
morphine sulphate+atropine sulphate	1140882116
morphine tartrate+cyclizine	1140882406
diamorphine	1140884460
methylmorphine	1140910376
diacetylmorphine	1140910402
oxycodone hydrochloride	1141171038
fentanyl+droperidol	1140879212
fentanyl	1140880956
fentanyl product	1141157470
pethidine	1140884388
methadone	1140884482
martindale methadone dtf 1mg/ml mixture	1140922628
heroin	1140888836
buprenorphine	1140871732
co-codamol	1140923346
co-proxamol	1140923348
co-dydramol	1140923350
co-codaprin	1140923344

Appendix S2. Methodology for exploratory and sensitivity analyses.

Exploratory analyses

Based on previous literature and clinical knowledge, we examined whether the association between chronic musculoskeletal pain and all-cause mortality differed by sex, age, BMI, ethnicity, or smoking status through testing of multiplicative interactions using WALD statistics (Harrell Jr, Frank E. *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*. Springer, 2015.). To avoid potential multiple testing issue, we chose the number of pain sites as the exposure and treated it as an unordered categorical variable.

Sensitivity analyses

1. To assess the influence of potential induction period, we used different lag time periods (1, 3, 5, and 7-year lag). Exposure status at a given time will correlate with a possible increase or decrease in disease only at some later time, which might introduce bias if we modelled the exposure-outcome association without considering this issue; lag period analysis could assess the potential influence by the induction period (Lash TL, VanderWeele TJ, Haneause S, Rothman K. *Modern epidemiology*. Lippincott Williams & Wilkins; 2020).

2. Except the covariates adjusted, there are other covariates which might be considered as confounders. However, the relationship between these covariates and the exposure might be bi-directional. Thus, we performed a sensitivity analysis including these covariates: body mass index (continuous), diabetes (yes or no), cancer (yes or no), depression (yes or no), anxiety (yes or no), cardiovascular disease (included heart attack, angina, stroke and high blood pressure; codes as the number of cardiovascular disease; the value ranged from 0-4). Two models were used: outcome regression and inverse probability treatment weighting through twang package. With twang package, gradient boosted models (number of tress was 5000 and 2000 for the analysis of pain type and number of pain sites, respectively) were used to calculate propensity score.

3. Severe cancer patients could have severe pain, which might bias the results. We could not identify severe cancer patients. Thus, we excluded participants with cancer.

4. To explore the potential influence from unmeasured confounding, E-value was calculated (VanderWeele TJ, Ding P. Sensitivity Analysis in Observational Research: Introducing the E-Value. *Annals of internal medicine.* 2017;167(4):268-274.).

In the main analyses, results from model 2 (analyses adjusted for age, sex, ethnicity, and the Townsend deprivation index) are reported in the results section. To increase readability for the above exploratory and sensitivity analyses, results from model 2 are presented in Appendix S4-S8.

Appendix S3. Results for exploratory analyses.

We found the association between chronic musculoskeletal pain and all-cause mortality differed by age. Considering the data distribution and clinical meaning, we chose 60 as the cut-off point. We also set 55 as another cut-off point to verify the results.

	Pain type				
	No pain	**Neck or shoulder pain**	**Back pain**	**Hip Pain**	**Knee Pain**
>= 60					
No of deaths (n=14159)	8670	1399	1625	593	1872
Multivariable adjusted[a]	1 (reference)	1.07 (1.01, 1.14)	1.16 (1.10, 1.23)	1.14 (1.05, 1.24)	1.01 (0.96, 1.06)
< 60					
No of deaths (n=5273)	3207	558	680	167	661
Multivariable adjusted[a]	1 (reference)	1.07 (0.98, 1.17)	1.16 (1.07, 1.26)	1.17 (1.00, 1.36)	1.10 (1.01, 1.19)
>= 55					
No of deaths (n=16801)	10281	1687	1941	687	2205

Multivariable adjusted[a]	1 (reference)	1.09 (1.03, 1.14)	1.17 (1.11, 1.22)	1.16 (1.07, 1.25)	1.02 (0.97, 1.07)
< 55					
No of deaths (n=2631)	1596	270	364	73	328
Multivariable adjusted[a]	1 (reference)	0.99 (0.87, 1.13)	1.15 (1.03, 1.29)	1.07 (0.85, 1.36)	1.14 (1.01, 1.28)
			Number of pain sites		
	No Pain	**One**	**Two**	**Three**	**Four**
>= 60 years					
No of deaths (n=18832)	8670	5489	2890	1310	473
Multivariable adjusted[a]	1 (reference)	1.08 (1.04, 1.12)	1.22 (1.17, 1.28)	1.36 (1.28, 1.44)	1.34 (1.22, 1.47)
< 60 years					
No of deaths (n=7085)	3207	2066	1059	537	216
Multivariable adjusted[a]	1 (reference)	1.11 (1.05, 1.18)	1.32 (1.23, 1.42)	1.65 (1.51, 1.81)	1.82 (1.59, 2.09)

>= 55 years					
No of deaths (n=22436)	10281	6520	3429	1599	607
Multivariable adjusted[a]	1 (reference)	1.09 (1.06, 1.12)	1.24 (1.19, 1.29)	1.40 (1.33, 1.48)	1.44 (1.33, 1.57)
< 55 years					
No of deaths (n=3481)	1596	1035	520	248	82
Multivariable adjusted[a]	1 (reference)	1.09 (1.01, 1.18)	1.33 (1.20, 1.47)	1.71 (1.49, 1.95)	1.60 (1.28, 2.00)

Data are presented as hazard ratio (95% confidence interval) unless otherwise indicated.

[a] Adjusted for age, sex, townsend deprivation index and ethnicity.

Appendix S4. Results for sensitivity analyses: induction period.

	Pain type				
	No pain	**Neck or shoulder pain**	**Back pain**	**Hip Pain**	**Knee Pain**
3 months					
No of deaths (n=19359)	11834	1949	2293	757	2526
Multivariable adjusted[a]	1 (reference)	1.07 (1.02, 1.13)	1.16 (1.11, 1.22)	1.15 (1.07, 1.24)	1.03 (0.99, 1.08)
6 months					
No of deaths (n=19221)	11746	1935	2275	752	2513
Multivariable adjusted[a]	1 (reference)	1.07 (1.02, 1.13)	1.16 (1.11, 1.22)	1.15 (1.07, 1.24)	1.04 (0.99, 1.08)
1 year					
No of deaths (n=18909)	11557	1894	2234	744	2480
Multivariable adjusted[a]	1 (reference)	1.07 (1.02, 1.12)	1.16 (1.11, 1.22)	1.16 (1.07, 1.25)	1.04 (0.99, 1.08)
3 years					
No of deaths (n=16966)	10395	1685	1989	665	2232

Multivariable adjusted[a]	1 (reference)	1.06 (1.00, 1.11)	1.15 (1.10, 1.21)	1.15 (1.06, 1.24)	1.04 (0.99, 1.09)
5 years					
No of deaths (n=14444)	8810	1447	1677	568	1942
Multivariable adjusted[a]	1 (reference)	1.07 (1.01, 1.13)	1.15 (1.09, 1.21)	1.16 (1.06, 1.26)	1.07 (1.02, 1.12)
7 years					
No of deaths (n=11204)	6805	1139	1299	447	1514
Multivariable adjusted[a]	1 (reference)	1.08 (1.02, 1.15)	1.15 (1.08, 1.22)	1.18 (1.07, 1.30)	1.08 (1.02, 1.14)
	Number of pain sites				
	No Pain	**One**	**Two**	**Three**	**Four**
3 months					
No of deaths (n=25817)	11834	7525	3929	1842	687
Multivariable adjusted[a]	1 (reference)	1.09 (1.06, 1.12)	1.25 (1.20, 1.30)	1.43 (1.37, 1.51)	1.46 (1.35, 1.58)
6 months					

No of deaths (n=25632)	11746	7475	3895	1832	684
Multivariable adjusted[a]	1 (reference)	1.09 (1.06, 1.12)	1.25 (1.20, 1.29)	1.44 (1.37, 1.51)	1.46 (1.35, 1.58)
1 year					
No of deaths (n=25216)	11557	7352	3825	1805	677
Multivariable adjusted[a]	1 (reference)	1.09 (1.06, 1.12)	1.24 (1.20, 1.29)	1.44 (1.37, 1.51)	1.47 (1.36, 1.59)
3 years					
No of deaths (n=22619)	10395	6571	3420	1619	614
Multivariable adjusted[a]	1 (reference)	1.08 (1.05, 1.12)	1.24 (1.19, 1.29)	1.43 (1.36, 1.51)	1.48 (1.37, 1.61)
5 years					
No of deaths (n=19280)	8810	5634	2900	1404	532
Multivariable adjusted[a]	1 (reference)	1.09 (1.06, 1.13)	1.24 (1.18, 1.29)	1.47 (1.39, 1.55)	1.52 (1.39, 1.66)
7 years					
No of deaths (n=14996)	6805	4399	2258	1119	415

Multivariable adjusted[a]	1 (reference)	1.10 (1.06, 1.15)	1.24 (1.18, 1.30)	1.51 (1.42, 1.61)	1.52 (1.38, 1.68)

Data are presented as hazard ratio (95% confidence interval) unless otherwise indicated.

[a] Adjusted for age, sex, townsend deprivation index and ethnicity.

Appendix S5. Results for sensitivity analyses: other covariates.

	Pain type				
	No pain	**Neck or shoulder pain**	**Back pain**	**Hip Pain**	**Knee Pain**
No of deaths (n=19441)	11877	1957	2305	769	2533
Outcome regression	1 (reference)	1.05 (1.00, 1.10)	1.12 (1.07, 1.17)	1.09 (1.01, 1.18)	0.97 (0.93, 1.02)
IPTW	1 (reference)	1.06 (1.01, 1.11)	1.10 (1.05, 1.15)	1.09 (1.00, 1.18)	0.99 (0.94, 1.04)
	Number of pain sites				
	No Pain	**One**	**Two**	**Three**	**Four**
No of deaths (n=25917)	11877	7555	3949	1847	689
Outcome regression	1 (reference)	1.05 (1.02, 1.08)	1.15 (1.11, 1.19)	1.25 (1.19, 1.32)	1.21 (1.11, 1.30)
IPTW	1 (reference)	1.05 (1.02, 1.08)	1.15 (1.11, 1.20)	1.29 (1.22, 1.36)	1.28 (1.16, 1.41)

IPTW: inverse probability treatment weighting. Data are presented as hazard ratio (95% confidence interval) unless otherwise indicated.

Included covariates in both models: age, sex, townsend deprivation index, ethnicity, body mass index, diabetes, cancer, depression, anxiety, and cardiovascular disease.

Appendix S6. Results for sensitivity analyses: exclude cancer patients.

	Pain type				
	No pain	**Neck or shoulder pain**	**Back pain**	**Hip Pain**	**Knee Pain**
No of deaths (n=16084)	9865	1598	1875	608	2138
Multivariable adjusted[a]	1 (reference)	1.06 (1.00, 1.11)	1.14 (1.09, 1.20)	1.13 (1.04, 1.23)	1.05 (1.00, 1.10)
	Number of pain sites				
	No Pain	**One**	**Two**	**Three**	**Four**
No of deaths (n=21395)	9865	6219	3221	1521	569
Multivariable adjusted[a]	1 (reference)	1.08 (1.05, 1.12)	1.24 (1.20, 1.30)	1.44 (1.37, 1.52)	1.49 (1.37, 1.63)

Data are presented as hazard ratio (95% confidence interval) unless otherwise indicated.

[a] Adjusted for age, sex, townsend deprivation index and ethnicity.

Appendix S7. Results for sensitivity analyses: E-value.

Pain type				
No pain	**Neck or shoulder pain**	**Back pain**	**Hip Pain**	**Knee Pain**
1 (reference)	1.34 (1.16)	1.62 (1.46)	1.57 (1.34)	1.21 (1)
Number of pain sites				
No Pain	**One**	**Two**	**Three**	**Four**
1 (reference)	1.4 (1.31)	1.81 (1.71)	2.21 (2.06)	2.28 (2.04)

Data are presented as hazard ratio (95% confidence interval) unless otherwise indicated.

Appendix S8. Transition probabilities.

	Neck or shoulder pain only	Back pain only	Hip pain only	Knee pain only	Number of pain sites (whole cohort)
Survival	0.9394385	0.9385606	0.9390286	0.9384451	0.9325723
Death due to cancer	0.03284171	0.03324539	0.03324311	0.03333191	0.03509146
Death due to cardiovascular disease	0.0114784	0.01161466	0.0116182	0.01175566	0.01306304
Death due to mental and behavioural disorder	0.001221386	0.001303996	0.001215858	0.001260145	0.001412712
Death due to respiratory system disease	0.003813001	0.003911987	0.003797145	0.003797522	0.004826117
Suicide	0.0005121942	0.0005285298	0.0005162573	0.0005040581	0.0005099293
Death due to nervous system disease	0.003322695	0.00333147	0.003333961	0.003314823	0.003611132
Death due to endocrine, nutritional and metabolic disease	0.0006260151	0.0005848485	0.0005645056	0.0005467749	0.0007674956
Death due to digestive system disease	0.001996244	0.00201881	0.001944408	0.002114481	0.002515825

Death due to musculoskeletal system and connective tissue disease	0.0001794868	0.0001949495	0.0001688692	0.0001922255	0.0002445579
Death due to genitourinary system disease	0.0003327073	0.0003292481	0.0003184391	0.0003289193	0.0004136671
Death due to fallls	0.0004290173	0.000472211	0.0004390599	0.0004442546	0.0004630991
Death due to other causes	0.003808623	0.003903322	0.003811619	0.003964118	0.004508712

CHAPTER EIGHT

Conclusions

Overview of findings

The first aim of this thesis was to evaluate the effectiveness and safety of conservative care approaches for pregnancy-related back pain. **Chapter Two** presented the results of a network meta-analysis that included 18 randomised controlled trials and 23 studies (randomised controlled trials and observational studies) in the qualitative synthesis. For women with back pain during pregnancy, progressive muscle relaxation therapy (mean difference [MD]: −3.96, Confidence interval [95% CI]: −7.19 to −0.74; moderate-quality evidence) and Kinesio Taping (MD: −3.71, 95% CI: −6.55 to −0.87; low-quality evidence) provided small reductions in pain intensity (Visual Analog Scale, range = 0 to 10) compared with placebo. Moderate-quality evidence suggested that transcutaneous electrical nerve stimulation improved physical function (MD: −6.33, 95% CI: −10.61 to −2.05; Roland Morris Disability Questionnaire, range = 0–24) compared with placebo.

The second aim of this thesis was to perform a network meta-analysis and systematic review to assess the efficacy and safety of surgical and invasive procedures for adults with degenerative lumbar spinal stenosis. While **Chapter Three** described the detailed protocol of the network meta-analysis, highlighting some of the methodological issues and clinical concerns in performing this study, **Chapter Four** presents the full results. The network meta-analysis included 49 randomised controlled trials with 5323 patients and 16 interventions. For the primary outcomes physical function and all-cause mortality, there were no statistically significant differences between any surgical or invasive intervention and conservative care. For several secondary outcomes (back pain, mobility, or treatment withdrawal due to any reason), no significant difference between groups was observed either. However, the review found that interspinous device (MD: -2.05, 95% CI: -3.98 to -0.12), midline splitting decompression (MD: -2.47, 95%CI: -4.45 to -0.5) and conventional open decompression (MD: -1.80, 95% CI: -3.49 to -0.11) were statistically superior to conservative care on short-term leg pain (0-10-point Visual Analog Scale) relief, although the differences were too small to be clinically important.

The third aim of this thesis was to examine both cross-sectional and longitudinal associations between lumbar radiographic changes and the severity of back pain-related disability. The results

in **Chapter Five** showed no evidence to support any association between higher number of lumbar segments with radiographic changes (K-L grade, osteophytes, and disc space narrowing) and more severe of back pain–related disability (e.g., cross-sectional analyses using the K-L grade; 1 segment vs 0 segment: adjusted odds ratio, 1.22, 95% CI, 0.76-1.96).

The fourth aim of this thesis was to identify distinct trajectories of analgesic use and the association of these trajectories with mortality and quality of life. The study presented in **Chapter Six** identified three distinct trajectories of analgesic use: (i) 'no use' group (691, 85.9%); (ii) 'increasing probability to use' group (73, 9.1%); and (iii) 'constant analgesic use' group (40, 5.0%). Compared with the 'no use' group, the 'constant analgesic use' group was associated with 2.15 times higher risk of all-cause mortality (95% CI, 1.18 to 3.91). There was no association between cause-specific mortality and pattern of analgesic use. Compared with 'no use' group, 'increasing probability to use' group and 'constant analgesic use' group were associated with worse quality of life in terms of physical function, role limitations due to physical health, and pain.

The fifth and final aim of this thesis was to quantify the association between chronic musculoskeletal pain and all-cause mortality and to investigate the extent to which this association was mediated by physical activity, smoking status, alcohol consumption, and opioid use. **Chapter Seven** showed that single pain sites in the neck or shoulder, lower back and hip were associated with higher risk of all-cause mortality. Higher number of pain sites was also associated with increased risk of all-cause mortality compared to having no pain (e.g., four sites vs no site of pain, Hazzard Ratio [HR] 1.46, 95% CI, 1.35 to 1.57). The single mediator analyses showed the following mediating proportions of the association between chronic musculoskeletal pain and all-cause mortality: 8.0% to 15.7% for physical activity; 32.5% to 79.0% for opioid use; 14.6% to 29.8% for smoking status and 2.4% to 17.5% for alcohol consumption. The multiple mediator analyses showed that the mediating proportion of all four mediators ranged from 53.4% to 122.6%: for participants with one pain site, chronic musculoskeletal pain was not associated with all-cause mortality; for participants with two or more pain sites, there was a reduction in the effect estimate,

for example, for two pain sites, HR reduced from 1.25 (95% CI 1.21 to 1.30) to 1.07 (95% CI 1.01 to 1.11).

Implications and directions of future research

Conservative care for pregnancy-related back pain

The review presented in **Chapter Two** helps to fill the gap in evidence regarding optimal treatment for pregnancy-related back pain. Clinicians have the evidence to support the use of progressive muscle relaxation therapy and Kinesio Taping to help decrease pain and the use of transcutaneous electrical nerve stimulation to help improve physical function. However, there are still several main issues which will require data from further high-quality trials before they can be addressed:

- Some interventions have only been investigated in one or two studies (e.g., progressive muscle relaxation therapy, Kinesio Taping, and transcutaneous electrical nerve stimulation)[1-3]. Further trials are needed to establish their efficacy.
- Some interventions (e.g., Kinesio Taping in combination with exercise and education) could not be integrated into the network meta-analyses given the lack of the essential studies connecting these interventions[4,5]. Therefore, the comparative effectiveness and safety of these interventions is still unknown.
- Future studies are needed considering the difference between pregnancy related back pain and pelvic girdle pain. A previous cohort study with 412 women in the Netherlands has established that these two types of pain have different prognosis[6]. For example, the study showed that women with pelvic girdle pain are more likely to develop limited mobility and need assistance eg wheelchair or crutches; than women with back pain only. However, this study did not assess the impact of symptom duration or intensity on the prognosis of these two types of pain, and future studies are still needed to better understand the differences between them. In addition, it is unclear whether these results could be extrapolated to other countries. Future studies are still needed to assess the effectiveness of Kinesio Taping and transcutaneous electrical nerve stimulation compared to usual care as quality of current evidence is low to moderate. Moreover, previous studies have shown the role of psychological factors (e.g., depression and fear avoidance) in low back pain. For example,

fear avoidance beliefs could mediate the relationship between pain and disability, result in poor physical health-related quality of life and increase health care utilisation in patients with low back pain[7]. But it is still unclear how to handle these psychological factors in pregnancy-related low back pain. Thus, an individualized way to assess and manage pregnancy-related low back pain might be needed. Future studies should also collect more accurate information about pain location, pain duration, pain intensity and different psychological factors and develop a clinical prediction model to recommend interventions based on these characteristics[8].

Surgical, invasive treatments and conservative care for degenerative lumbar spinal stenosis

The study presented in **Chapter Four** provides the most comprehensive and up to date evidence on the effectiveness and safety of treatments for degenerative lumbar spinal stenosis. These results may guide clinicians and consumers in their recommendations and choices for treatment. Previous clinical practice guideline development committees have had to rely on a small number of studies and two-intervention comparisons, resulting in conflicting recommendations. For example, the North American Spine Society clinical guidelines recommend epidural steroid injections to provide short-term (two weeks to six months) symptom relief in patients with lumbar spinal stenosis and associated neurogenic claudication[9]. This study however, showed that by assessing all available evidence, the recommendation of epidural steroid injection over conservative care cannot be endorsed. The new evidence available from this network meta-analysis, therefore, suggests that recommendations in current clinical practice guidelines for the management of degenerative lumbar spinal stenosis may need to be reconsidered. High quality cost-effectiveness analyses of surgical and invasive procedures for the management of degenerative lumbar spinal stenosis are also warranted to guide policy makers in their future recommendations. Of the 49 trials included in the network meta-analysis, only three included a cost-effectiveness analysis. Reduced walking capacity due to neurogenic claudication is an important outcome for degenerative lumbar spinal stenosis. However, the number of studies reported relevant outcomes is limited. Further studies should include this outcome. Future trials also need to report the details of their conservative treatment protocols more clearly in order to guide clinical practice. For instance,

while three trials included an exercise program, none provided enough detailed information on the type, dose, and duration of the programs to warrant replication in clinical practice.

Another important issue is the heterogeneity of patient presentation in degenerative lumbar spinal stenosis. Patients with different age, comorbidities, and disease severity might have different treatment responses. Several sensitivity analyses for the heterogeneities (e.g. level of stenosis, type of stenosis, difference in conservative care groups, typical symptoms and spinal instability) were performed to test the robustness of the results. In general, the results from sensitivity analyses were similar with the main analyses. But I acknowledge these might have been underpowered (e.g., 1-16 studies were excluded as different sensitivity analyses for short-term physical function which included 26 studies in the main analysis). Future studies should use individual patient data from multiple high-quality randomized controlled trials with extensive covariates to explore this issue through advanced statistical methods (e.g., The Predictive Approaches to Treatment effect Heterogeneity)[10]. Treatment responses might differ with different follow-up duration; thus, repeated measurements of covariates are preferred.

A final limitation concerns the definition of the term "degenerative spinal stenosis" as it is still unclear to set a clear age threshold. It is challenging to explore the role of participant age on the developmental of lumbar spinal stenosis in a meta-analysis based on aggregated data. Future studies should use individual-level data with sufficient sample size to explore the participant age issue in degenerative spinal stenosis[11].

Diagnostic imaging for lumbar spine

The results of **Chapter Five** may be used by clinicians and policy makers in educational campaigns for patients and the general public regarding the usefulness of lumbar radiographic findings. The lack of association between imaging findings and prognosis in terms of back pain–related disability, further adds to the evidence supporting the reduction of unnecessary imaging referrals. Future studies should include participants of both sexes and larger sample sizes and

should include multiple centres to increase external validity. Future studies may include data from L5/S1 segment and comprehensive positions of plain radiographs (anterior/posterior and lateral). With the data from L5/S1 segment, the assumption that lower levels of degenerative changes (i.e., L4/L5 and L5/S1) is associated with different symptoms could be tested. The results presented in **Chapter Five** however, do not support that view, as eight additional analyses (one exploratory analysis for potential interaction terms and seven sensitivity analyses for issues about the exposure definition, the potential population heterogeneity, the potential model misspecification, and the selection of confounder) have been performed with similar results to the main analysis. The association between the findings of complex imaging (e.g., computed tomography scans, magnetic resonance imaging, or nuclear bone scans) and symptom severity in people with back pain needs to be further explored, considering the increasing use of such imaging modalities. Studies including men in their sample, other age groups (not only middle-aged), other ethnic groups (e.g., Asian) or other countries (not only from the UK) should be performed.

Additionally, the potential high-risk subgroup should be explored. Although findings of diagnostic imaging might cause increased anxiety and use of care in some patients, some still believe it may be useful once the biology of back pain is better elucidated[12]. For example, people with similar lumbar spine radiographic changes might have different prognoses in terms of disability, based on their comorbidities (e.g., mental disorders, frailty index) they present with. The sub-population who response differently to the commonly used pharmacological or non-pharmacological treatments might have different prognoses. Thus, larger cohorts are needed considering the specific sub-population might have different prognoses.

Trends in analgesic use

From a policy perspective, it is important to know the clear association of analgesic use trajectories with mortality and quality of life. Presumably, a high and constant probability of using analgesics has even stronger mortality effects in recent years considering increasing opioid use with potential overdose. The findings (**Chapter Six**) indicate the public health initiatives aimed at addressing the potential drivers of a high and constant probability of using analgesics, including better

communication between physicians and patients and effective education for patients. The study did not present detailed information about the dosage and frequency of analgesic use, as this information was not available, which prevented us from exploring the potential dose-response relationship between analgesic consumption and mortality or quality of life. Overdose, or inappropriate choice of analgesic, should be explored in future studies. Similar to diagnostic imaging for lumbar spine, future studies looking at the impact of analgesic use should include both sexes, all age groups and a more ethnically and culturally diverse sample.

Due to the sample size limitation, detailed analyses considering specific type or kind of analgesic are not performed. Further studies should have sufficient sample size to explore this issue. For example, recent released clinical guideline recommended antidepressant medication for managing chronic pain[13]. However, the prognosis of the potential different pattern of antidepressants is unclear. Another recommendation for future studies is the study design. New-user design and target trial framework should be used, if possible, to reduce the potential biases[14]. Final issue is to identify potential drivers (e.g., central sensitization) for the analgesic use pattern. With better understanding of these drivers, we could locate high-risk sub-population (e.g., excessive use) easier and take relevant actions (e.g., education) to minimize potential harms.

Chronic musculoskeletal pain and mortality

The study in **Chapter Seven** comprehensively assesses the association between chronic musculoskeletal pain (type of pain and number of pain sites) and mortality (all-cause and cause-specific mortality). The results suggest that supporting healthy lifestyle behaviour (keeping adequate levels of physical activity, smoking cessation and alcohol consumption controlled) as well as opioids deprescription is an important strategy to decrease the mortality risk associated with chronic musculoskeletal pain. Several important issues should be explored in future studies:

- The role of detailed pain type is still unclear. For example, the widespread pain (e.g., fibromyalgia) versus regional pain and nociplastic versus neuropathic pain[15]. When further

data with detailed pain type as well as sufficient sample size and reasonable follow-up duration are available, new analyses on specific type of pain should be performed.

- Other pain characteristics should be explored. Patients with different pain intensity, pain duration, pain treatment responses and scale that pain interferes with daily life might be associated with various prognoses (e.g., all-cause mortality). In addition, patients with chronic musculoskeletal pain often report co-existing mental disorders (e.g., depression and anxiety). These psychological factors might also affect the prognosis. Future studies could separate the patients into different subgroups (e.g., high pain intensity vs moderate pain intensity vs minimal or no pain intensity) and then assess the prognosis of these subgroups.
- Repeated measurement of above variables might provide new perspectives. Patients' pain status (e.g., pain type, pain intensity, and pain duration) and relevant treatments could change with time. Thus, trajectory analysis capturing these changes could be used to separate patients into different subgroups and then assess the prognosis of these subgroups[16].
- Finally, cluster analysis by including these variables and relevant change patterns might be worthwhile to identify higher risk sub-population (e.g., more likely to die)[17].

Heterogeneity of musculoskeletal pain

This thesis includes two network meta-analyses (**Chapter Two and Four** with a research protocol at **Chapter Three**) and three cohort studies (**Chapter Five and Six**), which explores several aspects of the heterogeneity in musculoskeletal pain. **Chapter Two** focuses on pregnancy-related low back pain which is a distinct underlying health condition. People in this special life stage might respond differently to usual treatment strategies for non-specific low back pain. **Chapter Three and Four** focuses on degenerative lumbar spinal stenosis which is a different pain phenotype. Neurogenic claudication is a typical symptom for patients with degenerative lumbar spinal stenosis. Patients with this typical symptom might respond differently to treatments options compared with patients without. **Chapter Five** focuses on the radiological phenotype of lumbar spine. The lumbar spine radiographic change varies among patients with or without low back pain. These differences might indicate different prognoses (e.g., disability). **Chapter Six** focuses on the pattern of

analgesic use. Analgesics are commonly used in patients with musculoskeletal pain. Different analgesic use patterns might be associated with different prognoses (e.g., mortality and quality of life). **Chapter Seven** focuses on the mortality risk from chronic musculoskeletal pain. Different types of pain and number of pain sites might be associated with different mortality risk. In summary, considering the heterogeneity of musculoskeletal pain, the following issues were explored in this thesis: treatment strategies for pregnancy-related low back pain (a distinct underlying health condition), treatment strategies for degenerative lumbar spinal stenosis (a different pain phenotype), prognoses for lumbar spine radiographic changes (radiological phenotype of lumbar spine), prognoses for different analgesic use patterns, and mortality risk from chronic musculoskeletal pain with different pain types and number of pain sites. Definitively, there are many more issues to be explored considering their roles in the treatment strategies and prognoses for musculoskeletal pain, including but not limited to: psychological factors (e.g., depression), other chronic diseases (e.g., Alzheimer's disease), and other commonly used medications (e.g., antidiabetic drugs).

Concluding remarks

- For patients with back pain during pregnancy, progressive muscle relaxation therapy and Kinesio Taping may help to decrease pain, and transcutaneous electrical nerve stimulation may improve physical function, although the benefits might be perceived as being too small to be relevant to patients and clinicians.
- There was no evidence to support that surgery or invasive procedures are more effective or safer than conservative care in treating degenerative lumbar spinal stenosis.
- In a cohort of middle-aged, community-dwelling women, there was no evidence to support an association between a higher number of lumbar segments with radiographic changes (K-L grade, osteophytes, and disc space narrowing) and more severe back pain–related disability cross-sectionally or over time. These findings provide further evidence against routinely using diagnostic imaging of the lumbar spine.
- In a cohort of middle-aged women, a small group of participants had a high and constant probability of using analgesic over the study period and a markedly higher risk of all-cause mortality compared to those with no or low probability of using analgesics.

- At least half of the association of chronic musculoskeletal pain with increased all-cause mortality among middle-aged UK participants may be accounted for by a combination of four mediators (physical activity, smoking status, alcohol consumption, and opioid use). Our results suggest that supporting healthy lifestyle behaviour as well as reduced opioid use is an important strategy to decrease the mortality risk associated with chronic musculoskeletal pain.

References

1. Sedaghati P, Ziaee V, Ardjmand A. The effect of an ergometric training program on pregnants' weight gain and low back pain. *Gazzetta Medica Italiana*. 2007;166:209–213.
2. Akmeşe ZB, Oran NT. Effects of Progressive Muscle Relaxation Exercises Accompanied by Music on Low Back Pain and Quality of Life During Pregnancy. *Journal of midwifery & women's health.* 2014;59(5):503-509.
3. Kalinowski P, Krawulska A. Kinesio Taping vs. Placebo in Reducing Pregnancy-Related Low Back Pain: A Cross-Over Study. *Medical science monitor : international medical journal of experimental and clinical research.* 2017;23:6114-6120.
4. Mohamed EA, El-Shamy FF, Hamed H. Efficacy of kinesiotape on functional disability of women with postnatal back pain: A randomized controlled trial. *Journal of back and musculoskeletal rehabilitation.* 2018;31(1):205-210.
5. Pekçetin S, Özdinç S, Ata H, Can HB, Elter K. Effect of telephone-supported ergonomic education on pregnancy-related low back pain. *Women & health.* 2019;59(3):294-304.
6. Van De Pol G, Van Brummen HJ, Bruinse HW, Heintz AP, Van Der Vaart CH. Pregnancy-related pelvic girdle pain in the Netherlands. *Acta Obstet Gynecol Scand*. 2007;86(4):416-422.
7. Keeley P, Creed F, Tomenson B, Todd C, Borglin G, Dickens C. Psychosocial predictors of health-related quality of life and health service utilisation in people with chronic low back pain. *Pain*. 2008;135(1-2):142-150.
8. Chen L. Overview of clinical prediction models. *Ann Transl Med*. 2020;8(4):71.

9. Kreiner DS, Shaffer WO, Baisden JL, et al. An evidence-based clinical guideline for the diagnosis and treatment of degenerative lumbar spinal stenosis (update). *The spine journal : official journal of the North American Spine Society.* 2013;13(7):734-743.

10. Kent DM, Paulus JK, van Klaveren D, et al. The Predictive Approaches to Treatment effect Heterogeneity (PATH) Statement. *Ann Intern Med*. 2020;172(1):35-45.

11. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ*. 2010;340:c221.

12. Hall AM, Aubrey-Bassler K, Thorne B, Maher CG. Do not routinely offer imaging for uncomplicated low back pain. *BMJ*. 2021;372:n291.

13. Carville S, Constanti M, Kosky N, Stannard C, Wilkinson C; Guideline Committee. Chronic pain (primary and secondary) in over 16s: summary of NICE guidance. *BMJ*. 2021;373:n895.

14. Yoshida K, Solomon DH, Kim SC. Active-comparator design and new-user design in observational studies. *Nat Rev Rheumatol*. 2015;11(7):437-441.

15. Cohen SP, Vase L, Hooten WM. Chronic pain: an update on burden, best practices, and new advances. *Lancet*. 2021;397(10289):2082-2097.

16. Herle M, Micali N, Abdulkadir M, et al. Identifying typical trajectories in longitudinal data: modelling strategies and interpretations. *Eur J Epidemiol*. 2020;35(3):205-222.

17. Tibshirani, Robert, et al. *An introduction to statistical learning: With applications in R*. Springer, 2021.